Option Implied Risk Aversion under Transaction Costs: An Empirical Study

Siying Zhou

A Thesis

In

The John Molson School of Business

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Administration (Finance) at Concordia University

Montreal, Quebec, Canada

April 2018

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Siying Zhou

Entitled: Option Implied Risk Aversion under Transaction Costs: An Empirical Study

and submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN ADMINISTRATION (FINANCE)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

______________________________________Chair
Dr. Caroline Roux

______________________________________Examiner
Dr. Lorne Switzer

______________________________________Examiner
Dr. David Newton

______________________________________Supervisor
Dr. Stylianos Perrakis

Approved by ____
Graduate Program Director

____
Dean of Faculty

Date ______________April 18th, 2018_________________________________

Option Implied Risk Aversion under Transaction Costs: An Empirical Study

Siying Zhou

ABSTRACT

We empirically estimate the option implied coefficient of risk aversion of the market maker for European S&P 500 index options (SPX), involving asset allocation and option market making problems in the presence of proportional transaction costs in trading the underlying asset. We assume that the market maker has constant relative risk aversion utility and holds a two-asset portfolio consisting of the underlying and the riskless asset for a fixed, finite investment horizon which exceeds the option maturity, and she enters a position in the option market with an optimized portfolio. We follow the discrete time approach of Czerwonko and Perrakis (2016a, 2016b) to derive the market maker's simple investment policy and value functions, and apply a value matching condition to find option upper and lower bounds. Data on the S&P 500 index and the SPX options is collected over the period 1996–2016, 244 months in total, and the major variable, volatility, is re-estimated under the physical distribution. By matching observed SPX prices with numerically derived reservation prices, we estimate the level of implied risk aversion. Results show that in general, the market maker has lower risk aversion compared to investors who she trades with in order to accomplish a trade. A pattern that high risk aversion precedes rare market events is also exhibited, suggesting that a market maker may adopt a waiting policy if market events can be anticipated due to the information asymmetry.

ACKNOWLEDGMENT

First and foremost, I would like to express my sincere gratitude to my supervisor Dr. Stylianos Perrakis for taking me as his student, and for his unconditional help throughout my thesis. He has supported me with his patience, knowledge, and expertise. I have learned so much from him, and his kind advice will always guide me in the future. I also wish to thank Dr. Michal Czerwonko for his invaluable help with programming, modeling, and writing. He has always been there to answer my questions whenever I needed help. Without their assistance, this thesis would not have been written or completed.

I must express my true gratefulness to my parents and family for their unfailing support and continuous encouragement during my graduate studies. They have raised me with love and care, and they have always done their best to support me in my life. This accomplishment would not have been possible without them.

I also want to thank my partner Paxton, for his help in programming, thesis structure, as well as proofreading, and most importantly, for his consistent caring and support during the past two years.

Finally, I would like to thank my friends, both in Canada and in China, for their advice and support.

Table of Contents

List of Tables and Figures

List of Tables

List of Figures

Option Implied Risk Aversion under Transaction Costs: An Empirical Study

1. Introduction

This paper aims to empirically estimate the option implied coefficient of risk aversion (RRA) of the market maker for European S&P 500 index options (SPX). It is assumed that the market maker has unknown constant relative risk aversion utility (CRRA) and holds a two-asset portfolio consisting of the riskless asset and the underlying asset for a fixed, finite investment horizon, in the presence of proportional transaction costs in trading the underlying. The simple investment policy is numerically derived using the discrete time approach that converges to continuous time as the time partition tends to infinity, and option upper and lower bounds are derived using the value matching condition under a realistic assumption that the investment horizon exceeds the option expiration. The option implied risk aversion is empirically found by matching the numerically derived option prices to a wide range of different RRAs with the observed SPX bid and ask prices from January 1996 to April 2016, 244 months in total.

Index option pricing under transaction costs is a widely studied topic yet there is no satisfactory model. Merton (1989) first studied the option pricing for financial intermediaries who trade at the diametrically opposite sides of an investor. Early studies such as Leland (1985), Merton (1989), and Boyle and Vorst (1992) that are based on the no-arbitrage argument and portfolio replication only achieved trivial results. As alternatives, the stochastic dominance bounds for options with portfolio horizon exceeding the option maturity are valid for all risk averse investors and exclusively for univariate diffusion, as studied in Constantinides and Zariphopoulou (1999, 2001) and Constantinides and Perrakis (2002, 2007). The utility maximization approach first proposed by Hodges and Neuberger (1989) defines an investor's specific utility function, but its reservation option prices are found by assuming that the option maturity is the same as the portfolio horizon. Hence, these approaches are unsuitable for option pricing and market making problems of a market maker whose horizons differ. In this study, we present the first formal modeling of option market making which focuses on the fundamentals of option market. Our approach follows Czerwonko and Perrakis (2011) who applied discretization asset dynamics and asset allocation under transaction costs to option market making problems. Based on Constantinides' (1979, 1986) simple investment policy with respect to the two-asset portfolio selection problem, they formulated a model that numerically derives the European option bid and ask prices in the presence of proportional transaction costs from a market maker's perspective for a given risk aversion. This paper adopts the simple investment policy of Constantinides (1979, 1986) and follows the asset dynamics discretization of Czerwonko and Perrakis (2016a, 2016b) as well as their extension (Czerwonko and Perrakis, 2011) to estimate option upper and lower bounds and to investigate the level of risk aversion in the S&P 500 index option market from 1996 to 2016.

Market makers who work for financial institutions trade on the opposite side of the investors to ensure a certain level of liquidity in the option market. To successfully make the market, market makers need to predefine the optimal bid and ask prices to make sure they earn enough profits to compensate for bearing risks from price movements in the underlying asset and time decay of options. In this work, the market maker is assumed to hold the underlying and a riskless asset. She quotes reservation purchase and reservation write prices for an option whose maturity is earlier than her finite portfolio horizon. The reservation prices are defined as the price of the long or short options that make the market maker indifferent from adding or not options in her two-asset portfolio. The optimal ask (bid) price is the highest (lowest) price that she would be willing to sell (buy) the option, leaving her indifferent in terms of the derived utility of wealth, as discussed in Zakamouline (2006).

There are three steps to estimate the option implied risk aversion. First, we find the market maker's optimal investment policy. As shown originally by Constantinides (1979), in the presence of proportional transaction costs, the optimal revision of a two-asset portfolio follows a simple investment policy. That is, the market maker will not revise her portfolio when it lies in the convex-shaped no transaction (NT) region, otherwise she restructures her portfolio until it arrives to the nearest boundary of the NT region. The Euler discretization for continuous time diffusion is used to approximate the underlying asset dynamics in the market maker's two-asset portfolio selection problem without including options. Czerwonko and Perrakis (2016a, 2016b) have shown that their discrete time approach has computational advantages when the investment horizon is finite, and that it converges to the continuous time result as the time partition tends to infinity. For this discretization, we apply Kamrad and Ritchken (1991)'s trinomial lattice, which is suitable to approximate the logarithmic underlying return process. For our range of RRAs and a finite investment horizon, the simple investment policy and value functions are derived by solving an optimization problem whose objective is to maximize the market maker's expected utility of her terminal wealth. Second, we numerically derive the reservation purchase and reservation write prices for the market maker by altering her initial asset holdings for option prices and for portfolio revisions. We use a value matching condition under the assumption that the portfolio horizon exceeds the option expiration. The value function of holding options, which are dependent on option prices, are interpolated from the already derived value function in the two-asset portfolio case. This derivation of the value function with an option and subsequent interpolation are feasible thanks to the homogeneity of this value function of the same degree as the RRA, a property shown in Constantinides (1979) for CRRA utility. Thus, option prices that equate the value functions with or without holding options after asset holdings are adjusted for option exercise and portfolio revisions at the option maturity, become the estimates of reservation purchase and write prices. Last, we match the derived reservation prices with observed SPX prices for each observed option of a given moneyness over 244 observation dates. The Root-Mean-Square Error is used to measure the difference between two types of prices, and the Piecewise Cubic Hermite Interpolation is applied to find which RRA provides the best fit. For each observation date, the estimated RRA is the one whose corresponding derived prices have the best consistency with the market prices.

The results show that an increasing RRA shifts the boundaries of NT region toward the riskless asset, thus implying that a more risk averse market maker invests less in the risky asset. The mean of the estimated implied RRA of the market maker is 10.61, and it increases as the trading volume decreases, suggesting that a more risk averse market maker trades with a lower volume. In general, the market maker has lower risk aversion compared to investors, which implies that a market maker's lower (upper) bound needs to be higher (lower) than an investor's upper (lower) bound to accomplish a trade. Further, the estimated risk aversion is highly variable and peaks before rare market events. By plotting the estimated RRA of the market maker against observation dates, a pattern arises and shows that the estimated RRA is extremely high before a market event, and then drops significantly to a local minimum when the event starts, which suggests that should market events be anticipated, the trader adopts the waiting policy.

The rest of this work is organized as follows. Section 2 presents the literature review, and Section 3 presents the methodology. Section 4 describes the data and empirical results are discussed in Section 5. Conclusions and further research ideas are shown in Section 6.

2. Literature Review

Numerous studies have examined option pricing under transaction costs, under which the Black-Scholes (1973) model is invalidated by market frictions. In this section, we survey the main approaches: no-arbitrage and portfolio replication, stochastic dominance, and utility maximization. The two-asset portfolio selection problem in the presence of transaction costs in both continuous and discrete time is also described.

2.1 No-Arbitrage Argument and the Portfolio Replication

Early research on option pricing that considers transaction costs is based on the no-arbitrage argument, which attempts to replicate the payoff of derivatives by continuously rebalancing the portfolio independently of investor's risk preferences.

2.1.1 Modified Variance within the Black-Scholes

In a complete underlying market without transaction costs, a unique price of a European option can be derived using the Black-Scholes-Merton Model (BSM model), following the no-arbitrage argument which demonstrates that the price of a replicating portfolio consisting of N shares of stock and a riskless asset should be equal to the payoff of the option at any time. Under transaction costs, the original BSM model becomes invalid, since endless rebalancing results in infinite transaction costs.

Leland (1985) applied the modified variance ($\hat{\sigma}^2$) in the original BSM model. The payoff of call option $\hat{C}$ inclusive of percent transaction cost k is found by using the modified variance of an option inserted into the BSM model, which is defined as:

$$\begin{aligned} \hat{\sigma}^2(\sigma^2, k, \Delta t) &= \sigma^2[1 + kE\,|\frac{\Delta S}{S}|\,/\,\sigma^2 \Delta t] \\ &= \sigma^2[1 + \sqrt{(2/\pi)}k / \sigma\sqrt{\Delta t}] \end{aligned}, \qquad (2.1)$$

where $E\,|\frac{\Delta S}{S}| = \sqrt{(2/\pi)}\sigma\sqrt{\Delta t}$. The return of the underlying ($\frac{\Delta S}{S}$) is normally distributed with mean zero, and the rebalancing error would approach zero as $\Delta t \to 0$. The payoff $\hat{C}$ can be regarded as the upper bound of the option price, and a lower bound $\underline{C}$ can be found using the modified variance $\sigma^2 = \sigma^2 - kE\,|\frac{\Delta S}{S}|\,/\,\Delta t$ inserted into the Black-Scholes, implying that a portfolio replicates a short position in the option.

In Boyle and Vorst (1992), the self-financing portfolio that hedges a long call option at its maturity inclusive of the transaction costs has a value approximated by the BSM model with an adjusted variance: $\hat{\sigma}^2 = \sigma^2(1+\frac{2k\sqrt{n}}{\sigma\sqrt{T}})$, with $-k$ in the adjusted variance to find the payoff of a short call option. Because their adjusted variance is smaller than Leland (1985)'s by $\sqrt{(2/\pi)}$, their model derives a higher option value.

The modified variance is an intuitive approach to incorporate transaction costs into an option pricing model, yet it yields no satisfactory solutions because it confronts the same rebalancing dilemma as in the original BSM model. To decrease the rebalancing error, the difference between the option price and the value of the replicating portfolio at a time interval Δt, we need to increase the trading frequency. As $\Delta t \to 0$, the frequent portfolio rebalancing becomes extremely expensive no matter how small the transaction cost rate k is, and the modified variance tends to infinity (zero) for an upper (lower) bound.

2.1.2 Portfolio Replication and Super Replication

The binomial lattice approach of Cox, Ross, and Rubinstein (1979) provided a link between the approximation of underlying stochastic processes and the portfolio replication. Merton (1989) incorporated the proportional transaction costs in a two-period version of the Cox-Ross-Rubinstein (1979) binomial lattices and showed that the bid-ask spread can be substantial. He studied derivative security pricing from the financial intermediary's perspective. It is assumed that the financial intermediaries trade at the opposite side of an investor and hedge themselves perfectly by using the two-asset replicating portfolio. Later, Boyle and Vorst (BV, 1992) extended Merton (1989)'s model into multiple periods.

In a two-period binomial lattice, the underlying price S has the probability of u to go up and the probability of d to go down, and the option price is determined at each node of the binomial tree by the value of the replicating portfolio. Without transaction costs, the self-financing portfolio that contains N shares of stock and B dollars of riskless asset hedging a long call in one period is simply $NSu+RB$ ($NSd+RB$) if the stock price goes up (down). In the presence of transaction costs, the replicating costs $k\,|N-N_1|\,Su$ or $k\,|N-N_2|\,Sd$ need to be included in the portfolio value, under the transaction costs rate k ($k = k_1 = k_2$). Given that $N_2 \le N \le N_1$ for a portfolio replicating the long option, we must have the following at intermediate node of the binomial tree:

$$\begin{aligned} NSu(1+k) + BR &= N_1 Su(1+k) + B_1 R \\ NSd(1-k) + BR &= N_2 Sd(1-k) + B_2 R \end{aligned}, \tag{2.2}$$

which can be solved recursively from the call payoff at the option expiration date, at which the portfolios (N_1, B_1) and (N_2, B_2) are known. For the short option replication ($N_1 \le N \le N_2$), the relation becomes:

$$\begin{aligned} NSu(1-k) + BR &= N_1 Su(1-k) + B_1 R \\ NSd(1+k) + BR &= N_2 Sd(1+k) + B_2 R \end{aligned}. \tag{2.3}$$

However, the no-arbitrage argument imposes restrictions on the binomial model parameters, which need to follow the inequalities below to have meaningful hedging probabilities:

$$R(1+k) \le u(1-k),\ d(1+k) \le R(1-k), \tag{2.4}$$

which yields $\frac{1+k}{1-k} < \frac{u}{d}$. It will be violated when the number of periods of the binomial tree becomes large, because the right-hand sides gets smaller as the number of period increases.

Bensaid, Lesne, Pagès, and Scheinkman (BLPS, 1992) introduced the concept of super replication, which is an alternative to derive the option price in the framework of the binomial tree. Super replication can avoid infinite transaction costs caused by continuous trading since the portfolio need not be rebalanced at each node of the binomial tree. In the super replication model, the path is denoted by a set of ω and the transaction costs is a convex function. At any period t of the n periods to the option expiration ($t \in [0, n]$), we can define the stock price $S_t(\omega_t)$ and a replicating portfolio (N_t, B_t). There are two following paths ω_{t+1} with probability u and d for up and down moves, respecting to stock price uS_t and dS_t. BLPS (1992) introduced a dynamic algorithm whose objective is to minimize the initial cash position $(N_0 S_0 + B_0)$, which is required to cover the portfolio rebalance and transaction costs such that at each node of the binomial tree the portfolio satisfies the budget constraint: $RB_{t-1} > B_t + k(N_t(\omega) - N_{t-1}(\omega))S_t(\omega)$. For a cash-settled option, they showed that the derived upper bound is lower than or equal to that of the replicating strategy, which they interpret it as a potential saving on the transaction costs. Thus, they conclude that the super replication approach provides a tighter bound on option prices than the original replication approach.

Cases that are not discussed in Merton (1989), BV (1992), and BLPS (1992) were studied by Perrakis and Lefoll (1997). They introduced an algorithm that specifically derives option lower bound for the option with physical delivery. Their algorithm compliments the BLPS (1992) when the restriction in (2.4) is violated and when the underlying price lies within the interval $[K/(1+k), K/(1-k)]$ for portfolios that hedge a short option position. They also found that the derived lower bound of the European call options coincides with the Merton (1973) lower bound.

2.1.3 Problems of the Portfolio Replication

The no-arbitrage argument and portfolio replication approach did not provide a satisfactory solution for option pricing under transaction costs because their results become trivial even at realistic revision frequencies. The option bounds derived by Leland (1985) tend to the price of the underlying asset for a long option and the Merton (1973) lower bound $Max(0, S - Ke^{-rT})$ for the short option at the continuous time limit. As for the portfolio replication and super replication, the binomial parameters, $u = e^{\sigma\sqrt{\Delta t}}$, $d = u^{-1}$, and $R = e^{r\Delta t}$ would violate the restriction in (2.4) in the continuous time setting as $\Delta t \to 0$ with probability one. There is no solution for the case when this restriction is violated. As the number of periods tends to infinity, the initial position in the underlying converges to one and the position in the riskless asset tends to zero. An investor could simply establish an initial position in the underlying asset of one unit and pursue the buy and hold policy until the option maturity, instead of replicating continuously and wasting money on transaction costs. Davis and Clark (1994) conjectured, and Soner, Shreve, and Cvitanic (1995) proved that this trivial buy and hold strategy is the cheapest way to super replicate and hedge—a nontrivial hedging portfolio for option pricing with transaction costs does not exist.

2.2 Portfolio Selection under Transaction Cost

Without transaction costs, the optimal consumption and portfolio policy is defined by the "Merton line" (Merton, 1969), which is an optimal constant proportion of the risky to the riskless asset $\lambda^*/(1-\lambda^*)$, where $\lambda^* = (\mu - r)/[\sigma^2(1-\delta)]$. This quantity is only dependent on the risk premium, volatility of the underlying, and the risk aversion coefficient. However, when the transaction cost exists, asset allocation that simply follows the "Merton line" is no longer optimal since the required continuous rebalancing is prohibitively expensive. The simple investment policy, which was proposed and derived in a semi-closed form by Constantinides (1979, 1986), solved the portfolio selection problem and offered a path to study the option pricing problem under transaction costs. This approach is based on the capital equilibrium and has been researched in both continuous and discrete time.

2.2.1 Portfolio Selection Problem in the Continuous Time

Constantinides (1979) examined the effect of proportional transaction costs on portfolio selection problem based on early conjecture of Magill and Constantinides (1976). He further proved that the optimal investment policy of an investor who has a two-asset portfolio and isoelastic utility under the transaction costs is simple. This simple investment policy is described in terms of the no transaction (NT) region, which is shown to be a convex zone composed of two boundaries $[\underline{\lambda}, \overline{\lambda}]$ within which the investor is optimally refrained from trading. The investor would not trade if the proportion of the risky asset to the riskless asset λ_t $(\equiv y_t / x_t)$ in his portfolio lies within the NT region, i.e., $\lambda_t \in [\underline{\lambda}_t, \overline{\lambda}_t]$; otherwise, he would trade until the proportion λ_t arrives at the nearest NT boundary $\underline{\lambda}_t$ or $\overline{\lambda}_t$. Constantinides (1979) also showed that the value function $V(x_t, y_t, t)$ is monotone increasing and concave in (x, y) for an investor who holds x dollars in the riskless bond and y dollars in the risky asset. Constantinides (1986) derived the NT region $[\underline{\lambda}, \overline{\lambda}]$ in a semi-closed form, under the continuous time setting for an infinite horizon. The investor's consumption policy is assumed to be a constant at the rate β of the riskless asset ($c_t = \beta x_t$), and the asset dynamics within the NT region are:

$$\begin{aligned} dx_t &= rx_t - c_t dt \\ dy_t &= \mu y_t dt + \sigma y_t dw_t \end{aligned}, \tag{2.5}$$

where c_t is the consumption and w_t is a Wiener process. Then the author solves the optimization problem by maximizing the derived utility of the infinite consumption stream:

$$V[x_t, y_t; \beta, \underline{\lambda}, \overline{\lambda}] \equiv E_t \int_t^{\infty} e^{-\rho t} \Upsilon^{-1} c_t^{\Upsilon} dt \,, \tag{2.6}$$

where E_t is the expectation at time t, and ρ is the time discount factor. The relative risk aversion coefficient is assumed to be positive. If the proportion of the risky asset to the riskless asset λ_t lies within the NT region, the derived utility (2.6) satisfies the Bellman equation:

$$\frac{c^{\Upsilon}}{\Upsilon}+(rx-c)V_x+\mu yV_y+\frac{\sigma^2}{2}y^2V_{yy}-\rho V=0,\ \underline{\lambda}\le\lambda\le\bar{\lambda}\,. \tag{2.7}$$

When the proportion lies outside of the NT region, i.e., $\lambda_t \le \underline{\lambda}_t$ $(\lambda_t \ge \bar{\lambda}_t)$, trading is induced by purchasing (selling) v_t shares with changes $(1+k)v_t$ $((1-k)v_t)$ from the bond account, to reach the nearest NT boundary $\underline{\lambda}_t(\bar{\lambda}_t)$. In this case, the value function will satisfy the boundary conditions:

$$\begin{aligned}(1+k)V_x&=V_y,\ \lambda\le\underline{\lambda}\\(1-k)V_x&=V_y,\ \lambda\ge\bar{\lambda}\end{aligned}\,. \tag{2.8}$$

Combining with the boundary conditions in (2.8), the partial differential equation can be solved in a semi-closed form, that is, the value function $V(x,y,\beta)$ can be found by maximizing $V(x,y,\beta,\underline{\lambda},\bar{\lambda})$ with respect to the boundaries of NT region $[\underline{\lambda},\bar{\lambda}]$. Constantinides (1986) also found other properties of the NT region: the width of NT region goes up as the transaction costs increases, and transaction costs have more effect on the riskless account because the lower boundary $\underline{\lambda}$ decreases faster in k than the upper boundary $\bar{\lambda}$. The risk aversion coefficient δ and the variance σ^2 do not have the same effect on the width of the NT region, yet they shift NT region towards to riskless asset since the investors' demand for risky asset decreases as they get more risk-averse or as the variance increases. Constantinides (1986) also found the "liquidity premium" to be small, a quantity that he defined as the excess rate of return of the risky asset required to compensate the utility loss caused by transaction costs.

Davis and Norman (1990) studied the same subject as Constantinides (1986). They defined the proportional transaction costs as k_1 and k_2 for purchasing and selling the risky asset, respectively, which are incorporated into not only the boundary conditions but also the asset dynamics:

$$\begin{aligned}dx&=(rx-c)dt-(1+k_1)dL_t+(1-k_2)dU_t\\dy&=\mu ydt+\sigma ydw_t+dL_t-dU_t\end{aligned}\,, \tag{2.9}$$

where L_t and U_t are cumulative purchase and sale of the risky asset on the time interval $[0,t]$. Then they formed the optimization problem of maximizing the derived utility $E_0\int_0^{\infty} e^{-\rho t}u(c(t))dt$ for both isoelastic and lognormal utilities by stating a set of ordinary differential equations, which may be solved numerically. In the end, they arrived at the result that is qualitatively similar to Constantinides (1986). Their study further proved that the optimal investment policy is simple and the NT region is a wedge that reduces investor's trading.

Dumas and Luciano (1991) considered the same problem for power-utility investors who do not have intermediary consumption and maximize their derived utility of the terminal consumption at an uncertain future date, which tends to infinity. In their model, the control variables of the dynamic program are just the upper and the lower boundaries of NT region, $[\underline{\lambda}_t, \overline{\lambda}_t]$. In contrast with Constantinides (1986), they found that there was no shift of the NT region towards the riskless asset as transaction costs increase, which may be explained by a lack of changes in the riskless asset account caused by intermediate consumptions.

While Davis and Norman (1990) and Dumas and Luciano (1991) solved a free boundary problem for nonlinear differential equations, it is difficult to solve the Bellman equation in a semi-closed form when the investment horizon becomes finite, since this function becomes time-dependent. Liu and Lowenstein (2002) first solved the portfolio selection problem in a finite horizon in a continuous time setting. Under proportional transaction costs for sales of the risky asset, they examined the effect of an exponential (Erlang distributed) horizon on the investment policy of CRRA investors who maximize the derived utility of terminal wealth on a finite date. This date may be interpreted as their retirement date. The investor can purchase the risky asset at the price S_t or sell it at the price $(1-k)S_t$. The portfolio selection problem is then to choose the optimal amount of purchases and sales that maximizes the investor's derived utility of terminal consumption. The independent single Poisson event that governs this terminal date is exponentially distributed at time τ : $P\{\tau \in dt\} = \varphi e^{-\varphi t}dt$, where φ^{-1} is the investor's expected horizon parameter. For an investor with a long-life expectancy, the parameter φ is small. The constant finite horizon case corresponds to the convergence of the solution as the number of Erlang stages increases. Their model yields a semi-closed form solution for the value function determined numerically. They found that an investor with a short life expectancy may not invest in the risky asset that is subject to transaction costs at all even if the risk premium is positive, because the excess return of the risky asset may not be sufficient to cover the transaction costs.

2.2.2 Portfolio Selection Problem in the Discrete Time

The discrete-time two-asset portfolio section problem was first examined by Constantinides (1979). To purchase v_t shares of risky asset at date t, the investor needs to pay $(1+k_1)v_tS_t$ out of his bond account, whereas receives $(1-k_2)v_tS_t$ in the bond account when he sells, assuming no dividend payment of the risky asset. The asset dynamics in the form of discrete time are:

$$\begin{aligned} x_{t+1} &= [x_t - v_t - \max(k_1 v_t, -k_2 v_t)]R \\ y_{t+1} &= (y_t + v_t)\frac{S_{t+1}}{S_t} \end{aligned} \quad . \tag{2.10}$$

At each date t $(t \leq T-1)$, the investor maximizes his expected utility of terminal consumption in the form of the value function $V(x_t, y_t, t)$ of derived utility:

$$\max_v E_t\left[V\left([x_t - v - \max(k_1 v, -k_2 v)]R, (y_t + v)\frac{S_{t+1}}{S_t}, t+1\right)\right],\ t \leq T-1, \tag{2.11}$$

with the terminal condition $V(x_T, y_T, T) = U(x_T + y_T - \max[-k_1 y_T, k_2 y_T])$.

Genotte and Jung (1994) numerically derived the optimal trading strategies for two cases: the same transaction costs for both riskless and risky asset, and transaction costs only imposed to the risky asset. In their model, the risky asset dynamics follow the binomial lattice. Later, Boyle and Lin (1997) followed their methodology and presented numerical examples. However, there are possible numerical flaws in the discrete-time binomial model of Genotte and Jung (1994), and the authors acknowledged that their algorithm cannot handle option pricing problems under transaction costs.

Czerwonko and Perrakis (2016a) corrected the numerical flaws in Genotte and Jung (1994) and showed that the discrete time approach may provide a useful approximation of the continuous time approach when the time partitions tend to infinity in the fixed investment horizon. They applied the Euler discretization on dynamics of the risky asset for both diffusion and jump-diffusion cases, and presented an efficient numerical solution to the isoelastic utility investor's optimization problem for the derivation of the NT region, by applying the homogeneity of the value function first proven in Constantinides (1979). Compared to the continuous time approach in Liu and Lowenstein (2002), their study showed clearly that the discrete time approach has numerical computational advantages: it converges efficiently to the continuous time solution when the time partition becomes dense, and it outperforms the continuous time approximations by solving the finite horizon problem directly. It also admits a jump-diffusion process and other empirical features such as cash dividends.

Czerwonko and Perrakis (2016b) presented the flexibility of their discrete time approach in examining the economic impact on the portfolio selection problem related to several parameters of asset dynamics and of investor preferences. They relaxed the assumption of no dividends on the risky asset by adding them to the riskless asset account. They derived the investment policy under realistic transaction costs and dividend yields, and found that dividends have a very limited influence on the NT region. They also proved that the discrete time approach is flexible enough to solve the cases where the continuous time method fails. This work follows their discrete time approach to numerically derive the simple investment policy. Details on incorporating options into the investor's utility maximization problem are presented in the methodology section.

2.3 Stochastic Dominance Argument and Utility Maximization Approach

An alternative approach to finding the option bounds under transaction costs is the stochastic dominance argument, which applies to all risk-averse investors. Constantinides and Zariphopoulou (1999) incorporated proportional transaction costs into the stochastic dominance approach. They derived an upper bound on the reservation write price of a European call option when intermediate trading is allowed. Without assigning specific forms of investor utility function, they assumed it to be increasing and concave. In their model, the option maturity is equal to the portfolio horizon. Later they (Constantinides and Zariphopoulou, 2001) studied the problem for CRRA investors with multiple securities, including a riskless asset, the underlying stock, and derivatives. The risk aversion coefficient is restricted to be between zero and one, and investor cannot increase their expected utility by further trade. Under a more realistic assumption that the portfolio horizon exceeds the option expiration, Constantinides and Perrakis (2002) derived stochastic dominance bounds on reservation prices for European options for both single- and multi-period economy, and they extended the derivation to American options in their 2007 paper. Their model also has the advantage of accommodating a jump-process in the underlying price, and is independent of the initial portfolio position, provided that position contains a sufficient amount of underlying asset to guarantee solvency at option expiration time. The problem with stochastic dominance approach is a scarcity of numerical results due to computational issues.

Hodges and Neuberger (1989) first proposed to use the utility-based approach and the indifference argument to find reservation purchase and reservation write prices of an option. Specifically, the reservation purchase price is defined as the amount of money that makes the investor indifferent from holding or not the option in her portfolio in terms of the derived utility. Similarly, the reservation write price is the price at which investors will be indifferent between writing or not the option. They used a binomial lattice to calculate the reservation prices of European call options while leaving the convergence of their numerical algorithm unproven. Davis et al. (1993) applied the utility maximization approach to solve the European option pricing problem for exponential utility investors. Since Davis and Norman (1990) stated that it is difficult to numerically obtain the boundaries of the NT region for a portfolio with multiple risky assets, Davis et al. (1993) interpreted the risky asset as the market portfolio or an index. Based on the framework of Davis and Norman (1990), the derived boundaries of the NT region determine an investor's simple investment policy. The option write price is found by the value matching condition. They found a unique viscosity solution of the nonlinear partial differential equation that is satisfied by value functions derived from their arguments, assuming that the portfolio horizon is equal to the option expiration date, but they pointed out that this assumption made the interpretation of investors' reservation prices as market bid-ask prices unreasonable.

Zakamouline (2006) extended the work of Hodges and Neuberger (1989) and Davis et al. (1993) by including a component of the fixed transaction cost in his utility-based model. He also maintained the assumption that the portfolio horizon and the option expiration date are equal,[1] and studied the effect of transaction costs on the reservation option prices for constant absolute risk aversion (exponential) investors. He solved the discrete time numerical problem by applying the method of the Markov chain approximation for the case of European call options. He followed the

[1] This unrealistic assumption allows the computation of the optimal portfolio policy for the option writing or purchasing investor. To our knowledge, no one has derived that policy when the horizon exceeds option maturity.

indifference argument and applied it on the derived value functions: the value function without an option is $V(t,x,y)$ and the value function of holding an option is $J(t,x,y,S,\theta)$, where θ is the number of options the investor owns. Thus, based on the indifference argument, the reservation prices of θ European options are the prices that make the two value functions equal:

$$\begin{aligned} V(t,x,y) &= J^{b}(t, x-\theta P_{\theta}^{b}, y, S, \theta) \\ V(t,x,y) &= J^{w}(t, x+\theta P_{\theta}^{w}, y, S, \theta) \end{aligned}, \tag{2.12}$$

where P_{θ}^{b} is the highest price that the investor would pay to purchase options when he is indifferent from buying or not the options. P_{θ}^{w}, on the other hand, is the lowest price that the investor would write. Zakamouline (2006) also tried to explore the reason of empirical pricing bias such as volatility smiles, volatility term structure, and the bid-ask spread, but he found that these empirical pricing biases cannot be fully explained by only the transaction costs even if a component of fixed cost is included.

Czerwonko and Perrakis (2011)'s numerical approach corresponds to the expected utility maximization for a constant proportional risk aversion utility market maker whose portfolio horizon exceeds the option maturity. Allowing for jump components in the discrete time asset dynamics, they found that the number of options traded and risk aversion have a strong impact on the bid-ask spread, and these bid and ask quotes are tighter than the stochastic dominance bounds derived by Constantinides and Perrakis (2002).

2.4 Summary

To solve the portfolio selection problem under the transaction costs of the market maker, this paper follows the simple investment policy discussed by Constantinides (1986), and the discrete time approach presented by Czerwonko and Perrakis (2016a, 2016b). The utility maximization approach is preferred to the stochastic dominance approach since it allows estimating RRA from the observed market prices. The value matching condition (or the indifference argument) as well as the numerical model of Czerwonko and Perrakis (2011) are key ingredients in deriving option bounds under the assumption that the investment horizon exceeds the option maturity. Details of the methodology are presented in the next section.

3. Methodology

We consider the problem of a market maker who has CRRA utility and maximizes her expected utility of the terminal wealth within a fixed, finite investment horizon T. She holds only two assets in her portfolio, a riskless asset and a risky asset, with the natural interpretation of an index. She also trades on a cash-settled European option whose underlying is the risky asset in her portfolio, and the option expires before the end of her investment horizon. The problem we consider applies as well to an investor who follows the same two-asset portfolio policy as the market maker. We treat this case as an extension and provide details later on.

The formulation that we apply is justified for market makers by the fact that they are assumed to be agents of a financial institution or, to adopt the formulation of Shleifer and Vishny (1997), "highly specialized investors using other people's capital". These agents' performance is assumed to be monitored at the end of a fixed time interval, the horizon of our problem. Accordingly, our formulation is also applicable to market making in equity options. For the investors, on the other hand, the two-asset portfolio assumption can be justified only if the risky asset is an index, since indexing is a highly popular policy for a large class of investors.

Before entering a position in an option, the market maker solves a two-asset portfolio selection problem following the simple investment policy in the presence of the proportional transaction costs. She pays the proportional transaction costs at the rate of k when purchasing or selling the underlying, but not on trading in the riskless asset. While pursuing the same simple investment policy as in the two-asset case, she predefines the bid and ask prices of the European option before trading options with investors in the market. The option prices that she quotes are in the form of reservation purchase and reservation write prices, which make her indifferent from holding or not a given option in her portfolio.

The risk aversion of the market maker consistent with the observed market prices is estimated by comparing the option reservation prices derived under the above assumptions with the European SPX prices of a given moneyness and over 244 observation dates.

This section is organized as follows: first, the asset dynamics in both continuous time and discrete time are introduced. Then we present the derivation of the simple investment policy and option reservation prices. Last, we demonstrate the methodology of estimating the market maker's and investor's risk aversion coefficient.

3.1 Asset Dynamics in Continuous and Discrete Time

It is assumed that the market maker holds a two-asset portfolio, including x_t dollars in the riskless asset and y_t dollars in the underlying asset, which pays a dividend yield γ. The dynamics of the riskless asset are:

$$dx_t = rx_t dt + \gamma y_t dt \, , \tag{3.1}$$

where r is the continuously compounded risk-free rate and γ the dividend yield rate. The underlying's dynamics follow a univariate diffusion process:

$$dy_t = \mu y_t dt + \sigma y_t dw_t \, , \tag{3.2}$$

where μ and σ respectively are instantaneous ex-dividend mean and volatility parameters, and w_t is a standard Wiener process.

The Euler discretization for the continuous time diffusion process is used to solve the market maker's portfolio selection problem without the presence of an option. Czerwonko and Perrakis (2016a) proved that this discrete time approach converges to the continuous time limit as the time partition tends to infinity. Further, to solve this problem the dividend yield γ is added back to the index mean since Czerwonko and Perrakis (2016b) showed that the attribution of dividends between the risky and riskless assets accounts plays a limited role. For this reason, the underlying asset return is cum-dividend from now on.

Define the return of the risky asset as $Z_{t+\Delta t} = \dfrac{y_{t+\Delta t}}{y_t}$, which yields the following valid approximation of (3.2) in discrete time:

$$Z_{t+\Delta t} = 1 + \mu_t \Delta t + \sigma_t \varepsilon \sqrt{\Delta t} \, , \tag{3.3}$$

where ε is a random variable with mean zero and variance one, and $\Delta t = 1/252$ since we use daily trading frequency with 252 trading days in one calendar year. The trinomial lattice introduced by Kamrad and Ritcheken (1991) is applied to find the distribution of the return $Z_{t+\Delta t}$ ($\equiv e^{h(t)}$) when the portfolio lies within, below or above the NT region. Note that $h(t)$ is the approximating distribution of returns over the period $[t, t+\Delta t]$:

$$h(t)=\begin{cases}\phi & \text{with probability } p_1\\ 0 & \text{with probability } p_2\\ -\phi & \text{with probability } p_3\end{cases}, \tag{3.4}$$

where $\phi=\eta\sigma\sqrt{\Delta t}$, and $\eta \geq 1$. We have following probabilities p_1 , p_2 , and p_3 that correspond to these three returns above, respectively:

$$p(t)=\begin{cases}p_1=\dfrac{1}{2\eta^2}+\dfrac{(\mu-0.5\sigma^2)\sqrt{\Delta t}}{2\eta\sigma}\\ p_2=1-1/\eta^2\\ p_3=\dfrac{1}{2\eta^2}-\dfrac{(\mu-0.5\sigma^2)\sqrt{\Delta t}}{2\eta\sigma}\end{cases}. \tag{3.5}$$

Therefore, the dynamics of the riskless asset and the risky asset under the proportional transaction costs in the discrete time setting are:

$$\begin{aligned} x_{t+1}&=(x_t-v_t-k\,|\,v_t\,|)R\\ y_{t+1}&=(y_t+v_t)Z_{t+1}\end{aligned}, \tag{3.6}$$

where v_t is the optimal portfolio revision at time t , k represents the rate of the proportional transaction costs which is assumed to be the same for both purchases and sales, $k\,|\,v_t\,|$ represents the dollar amount of transaction costs by which the market maker changes her riskless asset holdings, and $R=e^{r\Delta t}\cong 1+r\Delta t+o(\Delta t)$. The underlying purchases (sales) are financed by sales (purchases) of $(1+k)v_t$ ($(1-k)v_t$) of the riskless asset.

3.2 Numerical Derivation of the Simple Investment Policy

The market maker faces the same optimization problem in both continuous time and discrete time settings, whose objective is to maximize her expected utility of the terminal wealth, which is converted to cash at the terminal date T :

$$\max_{v_t\ \ t\in[0,\mathrm{T}\text{-}1)} E_t[U(x_T+(1-k)y_T,\ T)], \tag{3.7}$$

where v_t is the time-t investment decision, and $U(.)$ presents the market maker's utility function, which is assumed to be the CRRA utility:

$$U(w_T) = w_T^{1-\delta} / (1-\delta), \tag{3.8}$$

where δ is the coefficient of relative risk aversion (RRA). Since Constantinides (1979) proved that the optimal investment policy under proportional transaction costs for an investor with a two-asset portfolio in the diffusion case is simple, the solution to this market maker's optimization problem determines the simple investment policy. This policy is summarized by the buy and sell boundaries of the NT region, denoted respectively as $\underline{\lambda}_t$ and $\overline{\lambda}_t$, and λ_t denotes the time-t risky to riskless asset proportion ($\equiv y_t / x_t$). We assume that the market maker solves her portfolio selection problem before she enters a position in options: she will not revise her portfolio in the presence of transaction costs as long as it lies within the convex-shaped NT region, but restructures her portfolio to the nearest sell or buy boundary if it drifts out of this region. We use the dynamic programming to solve for (3.7):

$$V(x_t, y_t, t) = \max_{v_t, t \in [0,..,T-1]} E_t[V(x_{t+1}, y_{t+1}, t+1)], \tag{3.9}$$

with the boundary condition:

$$V(x_T, y_T, T) = U(x_T + (1-k) y_T, T). \tag{3.10}$$

The value function specified in (3.9)–(3.10) is concave and homogenous of the degree $1-\delta$ as shown in Constantinides (1979).

To numerically find the boundaries of the NT region and the market maker's value function via the dynamic programming formulation (3.7)–(3.10), we follow the discrete time approach of Czerwonko and Perrakis (2016a). The numerical algorithm goes through the backward recursion to find the determinants of the simple investment policy and solves a one-period maximization problem in each recursive step. The only requirement is to know the value function one period ahead. This requirement is easily satisfied since at time $T-1$ the value function may be maximized in either boundary from the terminal condition (3.10) for a given one-period discrete probability space. Then the optimization procedure is repeated thus yielding a pair of boundaries at each time t. Because the value function is smooth and concave within the NT region, it may be easily interpolated, which yields this function virtually continuous—a key ingredient of our numerical work. If the portfolio moves outside the NT region, the value function may be easily found due to its homothetic property. The details of our numerical approach are presented in the following paragraphs.

Define the control variable b_t $(\equiv y_t / w_t)$ as the proportion of the underlying asset y_t to total wealth w_t $(\equiv x_t + y_t)$ at time t, and redefine the value function as $V(w_t, b_t, t)$. In principle, since adding nominal values of the riskless and risky asset accounts does not result in an economically meaningful quantity in the presence of transaction costs, w_t should rather be termed 'pseudo wealth'. For simplicity, we term it 'wealth' hereinafter. The control variable b_t implies the old control variable λ_t and is more numerically stable as argued in Czerwonko and Perrakis (2016a). At each time t ($t \in [0,...,T-1]$), the wealth w_t is set equal to \$1 and a value function $V(1, b_t, t)$ is derived. Hence, the dollar holdings in the riskless and underlying asset are $x_t = 1 - b_t$, and $y_t = b_t$, respectively. Before any portfolio revisions, the dynamics of the wealth w_t and the proportion b_t become:

$$\begin{aligned} w_{t+1} &= x_{t+1} + y_{t+1} = (1-b_t)R + b_t Z_{t+1} \\ b_{t+1} &= \frac{y_{t+1}}{w_{t+1}} = \frac{b_t Z_{t+1}}{(1-b_t)R + b_t Z_{t+1}} \end{aligned} . \tag{3.11}$$

We solve an optimization problem that maximizes expected utility of total wealth to derive the boundaries of the NT region, a convex zone characterized by the buy and sell boundaries, $\underline{b}_t$ and $\overline{b}_t$:

$$\begin{aligned} \underline{b}_t &= \arg\max_b \ E_t[(1+kb)^{1-\delta} V(w_{t+1}, b_{t+1}, t+1)] \\ \overline{b}_t &= \arg\max_b \ E_t[(1-kb)^{1-\delta} V(w_{t+1}, b_{t+1}, t+1)] \end{aligned} , \tag{3.12}$$

where the value function $V(w_{t+1}, b_{t+1}, t+1)$ is already known at time t. To derive the value function $V(w_{t+1}, b_{t+1}, t+1)$ that enters the expectations in (3.12), we equally space the already known time-

$t+1$ boundaries of the NT region $[\underline{b}_{t+1}, \overline{b}_{t+1}]$ to a suitable grid of points, and use each point in this set for interpolation. We will have three types of proportion b_{t+1} in dependence where this proportion falls relative to the NT region. We apply the simple investment policy at time-$t+1$ relative to the discretized wealth dynamics by verifying whether the NT boundaries $[\underline{b}_{t+1}, \overline{b}_{t+1}]$ would stipulate a trade. Specifically, the market maker would not revise her portfolio if the proportion b_{t+1} lies within the NT region ($b_{t+1} \in [\underline{b}_{t+1}, \overline{b}_{t+1}]$). If the proportion b_{t+1} drifts out of the NT region, portfolio revisions are necessary and her wealth will be adjusted in two ways: she would buy (sell) v_{t+1} shares of underlying to arrive the buy (sell) boundary $\underline{b}_{t+1}$ ($\overline{b}_{t+1}$). In the case that the proportion b_{t+1} is below the lower boundary $\underline{b}_{t+1}$, the market maker solves for the investment decision v_t to increase the proportion of the underlying asset to total wealth, where with the use of (3.6) we have $\dfrac{y_{t+1}+v_t}{w_{t+1}-kv_t} = \underline{b}_{t+1}$ yielding $v_t = \dfrac{\underline{b}_{t+1} - b_{t+1}}{1+k\underline{b}_{t+1}} w_{t+1}$, and by substituting the last quantity into $w_{t+1} - kv_t$, she would have a new level of wealth:

$$w'_{t+1} = \frac{1+kb_{t+1}}{1+k\underline{b}_{t+1}} w_{t+1}. \tag{3.13}$$

If the proportion b_{t+1} is above the upper boundary $\overline{b}_{t+1}$, the market maker decreases the proportion and solves for $\dfrac{y_{t+1}-v_t}{w_{t+1}-kv_t} = \overline{b}_{t+1}$, which in turn yields the new level of wealth:

$$w'_{t+1} = \frac{1-kb_{t+1}}{1-k\overline{b}_{t+1}} w_{t+1}. \tag{3.14}$$

The value function of wealth $V(1, b_{t+1}, t+1)$ also can be derived backward from the terminal condition (3.10) at each time $t, t \in [0, ..., T-1]$. Given the wealth adjusted for portfolio revisions in (3.13)–(3.14) and the derived value function $V(1, b_{t+1}, t+1)$ entering the expectations in (3.12), the value functions of the market maker without holding options, $V(w_{t+1}, b_{t+1}, t+1)$, become:

$$
\begin{cases}
(\frac{1+kb_{t+1}}{1+k\underline{b}_{t+1}} w_{t+1})^{1-\delta} V\left(1,\underline{b}_{t+1},t+1\right) & \text{for } b_{t+1} < \underline{b}_{t+1} \\
w_{t+1}^{1-\delta} V\left(1,b_{t+1},t+1\right) & \text{for } \underline{b}_{t+1} \le b_{t+1} \le \overline{b}_{t+1}\, , \\
(\frac{1-kb_{t+1}}{1-k\overline{b}_{t+1}} w_{t+1})^{1-\delta} V\left(1,\overline{b}_{t+1},t+1\right) & \text{for } b_{t+1} > \overline{b}_{t+1}
\end{cases}
\tag{3.15}
$$

where the simple investment policy is used to generate the new level of wealth and the homothetic property is applied to move the wealth level outside the value function. When the proportion b_{t+1} falls outside NT region, the value function is simply the known value of $V(1,b_{t+1},t+1)$ at either extreme of the NT region multiplied by a known quantity of the new level of wealth w'_{t+1}, as shown in (3.15). The value function is maximized recursively at all time epochs in (3.12) thus yielding the determinants of the simple investment policy for a given finite horizon T.

3.3 Derivation of Bounds on Option Prices

With a portfolio optimized for the two-asset case, the market maker trades on the European cash-settled option with a strike price K and expiration time T', set at one month for the empirical work. Denote the moneyness m_t $(\equiv K/S_t)$, equal to the strike price divided by the underlying price at trading date t. C_a and C_b are the ask and bid prices of the call option, while P_a and P_b are the prices of the put, respectively. We maintain the assumption that the same simple portfolio policy applies in the presence of options. This assumption is valid whenever the net position in options is relatively small with respect to total wealth. Otherwise, our derived bid-ask spread is wider and contains the market maker's optimal bid and ask prices, since our simple policy is suboptimal; hence, our results yield a potentially useful approximation.

To derive the option bounds, we consider a value matching condition in which the market maker's value function inclusive of an option equals the value function of two-asset portfolio:

$$
E_t\left[V\left(w_{T'}^e, b_{T'}, T\right)\right] = V(1, b_t^0, T)\,, \tag{3.16}
$$

where $w_{T'}^e$ is market maker's wealth after option exercise and portfolio revisions, and $b_{T'}$ is the corresponding proportion of the risky asset to total wealth at option maturity date. To derive the value functions in (3.16), the market maker's wealth is set equal to \$1 and the proportion of underlying asset to the total wealth b_t^0 is situated at the midpoint of the NT region before she adopts an option position, i.e., $b_t^0 = (\underline{b}_t + \overline{b}_t)/2$. The value function without holding an option in the right-hand side of (3.16) was derived in the previous two-asset portfolio case. The value function in the left-hand side of (3.16) for the option case is defined below.

3.3.1 Derivation of the Option Upper Bound

We first consider that the market maker writes n_c call options at time t for the total price c and for the notional value of the underlying index of \$1. After selling n_c call options, her wealth changes from \$1 to \$ $1+c$ with c attributed to the riskless asset account. The initial proportion b_t will also be adjusted to a new proportion b_i, $b_i = b_t/(1+c)$. Since this operation lowers the proportion of the underlying asset to total wealth, we consider the lower segment of the NT region $[\underline{b}_t, b_t^0]$ for the initial portfolio position and set the initial wealth equal to \$1. In our numerical scheme, we equally space this segment, which yields a set of b_i, $b_i \in [\underline{b}_t, b_t^0]$. This equally spaced segment implies that the following cash additions ($\equiv c_i$) bring the initial position of $1-b_t^0$ and b_t^0, in the riskless and underlying asset respectively, to a new proportion:

$$c_i = \frac{b_t^0 - b_i}{b_i}. \tag{3.17}$$

For each value of b_i in segment $[\underline{b}_t, b_t^0]$, we derive the market maker's riskless asset holdings $x_{T'}$ and underlying holdings $y_{T'}$ at the option maturity date T'. The initial asset holdings x_i ($\equiv 1-b_i$) and y_i ($\equiv b_i$) vary according to the dynamics:

$$\begin{matrix} x_{\tau+1} = x_\tau e^{r\Delta t} \\ y_{\tau+1} = y_\tau Z_{\tau+1} \end{matrix}, \quad \tau \in [t, T'], \tag{3.18}$$

where the underlying holdings follow the Kamrad and Ritcheken (1991) trinomial lattice as in (3.4)–(3.5). Since we utilize 30-day options and daily portfolio revisions, we consider this lattice for 21 daily periods, which correspond to the average number of trading days in a month. At each epoch τ ($\tau \in [t, T']$), from the trading date t to the maturity date T', we adjust asset holdings according to the simple investment policy in (3.13)–(3.14). At the option maturity, the asset holdings $x_{T'}$ and $y_{T'}$ inclusive of the initial cash received for a written call need to be adjusted for the option exercise for each c_i. Since these holdings correspond to the total wealth of \$1 as stated above, we also need to multiply both holdings at the maturity by a factor of $1 + c_i$.

The call option is cash settled and with the payoff ($\equiv P_c$), $P_c = n_c S_t \left[Z_{T'} / (1+\gamma) - m_t \right]^+$, where $Z_{T'}$ is the cumulative return on the risky asset and the superscript $^+$ denotes the maximum of a given quantity and zero. For simplicity, the initial value of the underlying S_t is set to \$1. The settlement of the option payoff is divided into three situations in relation to sufficiency of the riskless asset holdings $x_{T'}$ for the option payoff P_c:

(1). $x_{T'} \geq P_c$, the market maker has sufficient cash holdings to settle the option payoff. Her riskless and risky asset holdings adjusted for the option exercise before any portfolio revisions are:

$$\begin{aligned} x_{T'}^e &= (1+c_i) x_{T'} - n_c \left(\frac{Z_{T'}}{1+\gamma} - m_t \right)^+ \\ y_{T'}^e &= (1+c_i) y_{T'} \end{aligned} . \tag{3.19}$$

(2). $x_{T'} < P_c$, the market maker does not have sufficient cash holdings and she needs to sell the equivalent number of shares of the underlying to settle her position. In this case, her asset holdings become:

$$\begin{aligned} x_{T'}^e &= 0 \\ y_{T'}^e &= (1+c_i) y_{T'} - [n_c \left(\frac{Z_{T'}}{1+\gamma} - m_t \right)^+ - (1+c_i) x_{T'}] / (1-k) \end{aligned} . \tag{3.20}$$

(3). $x_{T'} \leq 0$, when borrowing is optimal the option exercise only affects her risky asset holdings:

$$y_{T'}^e = (1+c_i) y_{T'} - [n_c \left(\frac{Z_{T'}}{1+\gamma} - m_t \right)^+] / (1-k) . \tag{3.21}$$

After the asset holdings are adjusted for the option exercise, they undergo the portfolio revision (3.13)–(3.14) due to the simple investment policy. After this step, we derive a set of value functions $E_t\left[V\left(w^e_{T'}, b_{T'}, T\right)\right]$ that corresponds to the set of candidate call prices c_i in (3.17). Thus, we arrive at a tabulated set of value functions, which can be easily interpolated by Hermite polynomials from the time-T' value functions derived in the two-asset portfolio case in (3.15). By using the value matching condition (3.16) and the fact that we may easily interpolate the value function at the option expiration, we can precisely estimate the call upper bound ($\equiv \overline{c}$) by a simple univariate optimization routine:

$$\overline{c} = \arg\min_{c_i} \left\{E_t\left[V\left(w^e_{T'}, b_{T'}, T\right)\right] - V(1, b^0_t, T)\right\}^2, \tag{3.22}$$

where the function $V\left(w^e_{T'}, b_{T'}, T\right)$ is derived numerically via interpolation. The interpretation of (3.22) is that it searches for a value of n_c call options that exactly satisfies the value matching condition in (3.16); the satisfaction of this condition may be easily verified.

If adding cash c to the riskless asset account from writing call options pushes the proportion of the underlying to riskless asset outside the NT region, i.e., $b_i < \underline{b}_t$, the initial asset holdings for the riskless asset and the underlying asset are simply $1 - \underline{b}_t$ and $\underline{b}_t$, since the market maker follows the simple investment policy and will adjust her wealth to the nearest boundary $\underline{b}_t$ at trading time t. This implies that both asset holdings at expiration time T', $x_{T'}$ and $y_{T'}$, need to be multiplied by a new factor of $\dfrac{1 + kb^0_t + c_j}{1 + k\underline{b}_t}$ instead of $1 + c_i$ in (3.19)–(3.21) before we adjust them for option exercise and portfolio revisions. Another problem to consider is the market maker's solvency condition. If the total wealth at the option maturity is negative at some option exercise value, the time-t expectations of the value function in the presence of a written call are set equal to $-\infty$.

Last, to derive the reservation write price of the call option per \$1 of the index or the call ask price c_a, we adjust the call upper bound $\overline{c}$ for the depth of the quote:

$$c_a = \overline{c} / n_c. \tag{3.23}$$

To derive the reservation write price of a put option, we simply replace the call option payoff P_c with the put option payoff $P_p = n_p \left[m_t - Z_{T'} / (1+\gamma) \right]^+$ in (3.19)–(3.21). We can derive a put upper bound $\overline{p}$ with the same procedure as above, and the put ask price is its upper bound adjusted for the depth of the quote: $p_a = \overline{p} / n_p$.

3.3.2 Derivation of the Option Lower Bound

Since this case is similar to the derivation of the upper bound in many aspects, our presentation mainly focuses on differences between them. The derivation of the lower bound is in fact simpler, since the solvency condition only applies at the option purchase and the option payoff is attributed to the riskless asset account. In this case, the market maker receives the option payoff at its maturity and pays cash up front. For buying n_c European call option at time t, the market maker pays c to an investor who sells n_c options, and thus her initial wealth changes from \$1 to \$ $1-c$. Hence, the initial proportion b_t is adjusted to the new proportion b_i, $b_i = b_t / (1-c)$, for the initial wealth set equal to \$1. Since this operation increases the proportion of the underlying to total wealth, we now consider the equally spaced upper segment of the NT region $[b_t^0, \overline{b}_t]$. The candidate call prices c_i implied by this segmentation, which now are cash subtractions, become:

$$c_i = \frac{b_i - b_t^0}{b_i}. \tag{3.24}$$

For each call price c_i, we apply again the initial asset holdings through the trinomial dynamics (3.18) until the option maturity. Since now the initial wealth is \$$1-c$ instead of \$1, the holdings at the option maturity, $x_{T'}$ and $y_{T'}$, are multiplied by a factor of $1-c_i$. Upon the option exercise, these asset holdings become:

$$\begin{aligned} x_{T'}^e &= (1-c_i) x_{T'} + n_c \left(\frac{Z_{T'}}{1+\gamma} - m_t \right)^+ \\ y_{T'}^e &= (1-c_i) y_{T'} \end{aligned}. \tag{3.25}$$

The rest of the derivation of the lower bound $\underline{c}$ is similar to the derivation of the upper bound. We adjust the asset holdings for portfolio revisions and apply the value matching condition. If the cash outlay for purchasing options pushes the underlying to riskless asset proportion out of the NT region, i.e. $b_i > \bar{b}_t$, both asset holdings at time T', $x_{T'}$ and $y_{T'}$, are multiplied by the new factor $\frac{1-kb_t^0+c_j}{1-k\bar{b}_t}$ instead of $1-c_i$ in (3.25). Finally, to arrive at the reservation purchase price or the call bid price c_b, we adjust it for the depth of the quote, namely, $c_b = \underline{c}/n_c$. For put options, we use the payoff $P_p = n_p\left[m_t - Z_{T'}/(1+\gamma)\right]^+$ in (3.25), and adjust the put lower bound for the depth of the quote, which yields $p_b = \underline{p}/n_p$.

3.4 Estimation of the Risk Aversion

To find the option implied risk aversion, we first derive option reservation prices that depend on the RRA. Specifically, we derive those prices for 15 different candidate RRAs ranging from 0.5 to 30.5. Our estimates of the implied RRA will be characterized by the best match with a given cross-section of observed market option prices as explained below. We assume that a single market maker quotes a call or put bid or ask price for all options in a given cross section. This assumption is not justified, neither economically nor numerically, since there is no reason for the same market maker to quote the lowest bid or ask price for different degrees of moneyness and for call and put options; further, the degree of approximation of our numerical approach is not independent of moneyness. It is also not necessary, since each observed quote can be matched individually with the corresponding estimate to determine the implied RRA that generally differs by the degree of moneyness. It does, however, allow us to represent with a single number the evolution of the quote-implied RRA across time, which uncovers some interesting empirical facts. A single implied RRA will be derived in the base case as the "best match" for all option quotes, both calls and puts and both long and short, in a given cross section.

Piecewise Cubic Hermite Interpolation is used to create a structure that contains 15 option prices derived for a given moneyness and for each RRA in our range. This structure separately includes upper and lower bounds of call or put options. Since these bounds are derived for the notional value of the underlying index of $1, now it suffices to multiply them by the observed value of the index to make them comparable to the market prices. We exemplify our structure with a matrix that corresponds to the call upper bounds:

$$\begin{bmatrix} c_{a_{m_1,\delta_1}} & c_{a_{m_1,\delta_2}} & \cdots & {}_{m_1,\delta_{15}} \\ c_{a_{m_2,\delta_1}} & c_{a_{m_2,\delta_2}} & \cdots & \\ \vdots & \vdots & \vdots & \vdots \\ c_{a_{m_n,\delta_1}} & c_{a_{m_n,\delta_2}} & \cdots & \end{bmatrix} = [c \qquad c_{a,2}, \quad \dots, \quad c_{a,n}]', \qquad (3.26)$$

where m and δ present the moneyness and RRA respectively, and n is the number of options observed at a given observation date. This structure is easy to interpolate since the sole variable of interpolation is RRA.

We minimize the Root-Mean-Square Error (RMSE) to find the best fit of the derived reservation prices to the observed SPX prices. The estimated RRA $\hat{\delta}$ is the one whose corresponding derived prices are closest to the market prices in terms of RMSE, found in the base case from:

$$\hat{\delta}_{Mm} = \arg\min_{\delta}\left(\sqrt{\frac{f}{2(n_1+n_2)}}\right), \qquad (3.27)$$

$$f = \sum_{m=1}^{n_1}\left(\frac{C_{a,m}(\hat{\delta}) - C_{a,m}}{C_{a,m}}\right)^2 + \sum_{j=1}^{n_1}\left(\frac{C_{b,j}(\hat{\delta}) - C_{b,j}}{C_{b,j}}\right)^2 + \sum_{k=1}^{n_2}\left(\frac{P_{a,k}(\hat{\delta}) - P_{a,k}}{P_{a,k}}\right)^2 + \sum_{h=1}^{n_2}\left(\frac{P_{b,h}(\hat{\delta}) - P_{b,h}}{P_{b,h}}\right)^2$$

where n_1 is the number of observed call options, n_2 is the number of the observed put options at a given date. $C_{a,m}$, $C_{b,j}$, $P_{a,k}$, and $P_{b,h}$ denote the observed market prices for different moneyness. $C_{a,m}(\hat{\delta})$, $C_{b,j}(\hat{\delta})$, $P_{a,k}(\hat{\delta})$, and $P_{b,h}(\hat{\delta})$ respectively denote the corresponding prices that are interpolated from evaluation of the piecewise polynomial in the already derived reservation prices for a given RRA.

As already noted, we may inverse the problem and look at the reservation write and purchase prices of an investor, which are the opposite relative to the ones for the market maker, in other words, the market maker lower bound on an option becomes the investor's upper bound and vice versa. We find the investor's RRA by simply matching derived option prices with observed market prices reversely, namely, we match the derived reservation purchase (write) prices with the observed market bid (ask) prices. Analogously to the derivation for the market maker, the estimated RRA of the investor for the base case is:

$$\hat{\delta}_{inv} = \arg\min_{\delta}\left(\sqrt{\frac{f}{2(n_1+n_2)}}\right). \tag{3.28}$$

$$f = \sum_{m=1}^{n_1}\left(\frac{C_{a,m}(\hat{\delta}) - C_{b,j}}{C_{b,j}}\right)^2 + \sum_{j=1}^{n_1}\left(\frac{C_{b,j}(\hat{\delta}) - C_{a,m}}{C_{a,m}}\right)^2 + \sum_{k=1}^{n_2}\left(\frac{P_{a,k}(\hat{\delta}) - P_{b,h}}{P_{b,h}}\right)^2 + \sum_{h=1}^{n_2}\left(\frac{P_{b,h}(\hat{\delta}) - P_{a,k}}{P_{a,k}}\right)^2$$

4. Data

Our data encompasses the period from January 1996 to April 2016. Underlying index data set consists of the prices of the S&P 500 index, its dividends and volatility, as well as the VIX index. The index option data set consists of strike prices, bid and ask prices, and time to maturity of SPX. The riskless rate is proxied by the three-month interest rate of the secondary market T-bill. As explained below, the volatility is estimated under the physical (P)-distribution.

4.1 Estimated Volatility

Two different methods are used to estimate the volatility under the physical or P-distribution. The first method uses the VIX index adjusted for an average forecasting error (AFE). The second method uses a linear regression forecasting model. By comparing the prediction error of each method, defined as the value of the estimated volatility ($\hat{\sigma}_t$) less the realized volatility (RV_t), we choose the method that generates the lowest prediction error. In either case, our estimate for the volatility contains only information available to traders at a given date.

In the first method, the daily VIX index is collected from the Chicago Board Options Exchange (CBOE) for the 1990–2016 period, and the initial volatility forecast is the value of VIX at the last day τ of calendar month t: $VIX_t = VIX_\tau$. Note that since our options data starts in 1996, we have at least six years in sample used to estimate the volatility. The realized volatility for month t is calculated as

$$RV_t = \sqrt{12\sum_{i=\tau+1}^{\tau+30}(\ln S_i - \ln S_{i-1})^2}\,, \tag{4.1}$$

where S_i is the price on trading day i and S_{i-1} is the price on the previous trading day $i-1$ and $\ln$ denotes the natural logarithm. Hence, the cumulated AFE is the average monthly difference between the implied and realized volatilities:

$$AFE_t = \frac{1}{n}\sum_{i=1}^{n}(VIX_t - RV_t) \ , \tag{4.2}$$

where n is the number of months cumulated from January 1990 to the observation month t. Then, the monthly estimated volatility $\hat{\sigma}_t$ is each value of the VIX at the observation date minus the cumulated AFE:

$$\hat{\sigma}_{t,1} = VIX_{obs} - AFE_t \, . \tag{4.3}$$

In the second method, the monthly realized and implied volatility, RV_t and VIX_t are calculated in the logarithmic form and used in linear regressions. Using the logarithm assures that our volatility estimate is always positive. Two linear regression models are used, with and without the intercept:

$$\begin{aligned} \log(RV_t) &= \alpha + \beta\log(VIX_t) + \varepsilon_t \\ \log(RV_t) &= \beta\log(VIX_t) + \varepsilon_t \end{aligned} \, . \tag{4.4}$$

By inverting the estimates from the regression above, the forecasted volatilities $\hat{\sigma}_t$ respectively are:

$$\begin{aligned} \hat{\sigma}_{t,2} &= \exp(\hat{\alpha} + \hat{\beta}\log(VIX_t)) \\ \hat{\sigma}_{t,3} &= \exp(\hat{\beta}\log(VIX_t)) \end{aligned} \, . \tag{4.5}$$

Table 1 presents the statistics for the prediction error of each method. Since our first method that uses the VIX subtracting the AFE has the lowest mean of the prediction error, it is selected as the predictor of the volatility under the *P*-distribution.

[Insert Table 1 here]

4.2 Index and Options Data

The historical daily prices of the S&P 500 index are collected from CBOE, and the daily strike prices, highest closing bid price, and lowest closing ask price of the S&P 500 index options (SPX) are collected from Option Metrics from January 1996 to April 2016, 244 months in total.

The exercise value of European style SPX options is determined at the market opening on the third Friday in each month before it expires on the following Saturday. Our target observation date is 30 days before the expiration date of each month. If there is no option traded on the target observation date, the nearest trading date before the target date is used instead. For each strike price K, the corresponding moneyness of the option is calculated as $m_t = K / S_t$, where S_t is the price of the underlying at the observation date t. We select call options that have original bid price no less than \$0.15 and moneyness between 0.96 to 1.08, as well as put options that have original bid price no less than \$0.15 and moneyness below 1.04.

The three-month secondary market T-bill daily interest rate (i_t) from 1996 to 2016 is collected from the Federal Reserve Economic Data. Since it uses the actual/360 calculation, the daily interest rates are adjusted to the annual rate as $i = 365 i_t / 360$. The annual dividends of S&P 500 Index are collected from the Robert Shiller online data, which is adjusted as the dividend yield for a given month using $Div_m = Div_{annual} / (12 \times S_t)$. Table 2 presents descriptive statistics of variables for the estimated volatility, interest rate, dividends, S&P 500 index price, bid-ask prices, strike prices, and moneyness of SPX.

[Insert Table 2 here]

5. Empirical Results and Discussion

5.1 Derived NT Boundaries and Option Bounds

The following set of parameters is used for deriving the boundaries of NT region: proportional transaction costs rate k of 0.25%, risk premium of 4%, investment horizon T of 10 years, 244 re-estimated volatilities and interest rates, and 15 different coefficients of risk aversion (δ) ranging from 0.5 to 30.5.

Table 3 presents the derived NT region for fixed investment horizon T = 10 years, and Table 4 shows the derived mid-point of the NT region $b^0_{T'}$ $(\equiv (\underline{b}_{T'} + \overline{b}_{T'})/2)$ at option maturity T', for 15 RRAs, and for 6 different volatilities.

[Insert Table 3 and Table 4 here]

As the results show, the risk aversion has a strong effect on the market maker's investment policy. The NT boundaries and their midpoint decrease as the RRA increases, implying that a more risk averse market maker would invest less in the risky asset. The same relation between the volatility and the boundaries holds, suggesting that the market maker would invest less in the risky investment when the market is more volatile. It is consistent with the properties of the NT region discussed in Constantinides (1986) which states that an increase in risk aversion or an increase in volatility shifts the NT region toward the riskless asset.

Once we have the derived boundaries and the value function at each observation date, we may derive option prices for a given depth of the quote, arbitrarily chosen at n_c of 0.2 or of 0.1.[2] A sample of the proportional bid-ask spreads, defined as the ratio of the difference between the upper and lower bound to their mean value, are presented in Tables 5 and 6.

[Insert Table 5 and Table 6 here]

In Table 5, the proportional bid-ask spreads are calculated for call options as $\frac{2(c_a - c_b)}{(c_a + c_b)}$. We use four degrees of moneyness 0.975, 1, 1.025, and 1.05, representing the in-the-money (ITM), at-the-money (ATM), out-of-the-money (OTM), and far out-of-the-money (far OTM) call options respectively. For put options, we calculate this quantity as $\frac{2(p_a - p_b)}{(p_a + p_b)}$ in Table 6, and four degrees of moneyness of 0.95, 0.975, 1 and 1.025 are selected, for far OTM, OTM, ATM, and ITM puts, respectively. Other parameters used in the algorithm that derives the option prices are riskless rate of 2.32%, estimated volatility of 16.06%, and dividend yield of 0.15%, representing the mean values of these variables in our data set.

[2] We assume implicitly that the market maker does not change her option position during the life of the option. The bounds continue to be valid under such an assumption.

We can observe the effect of the RRA on the market maker's quotes from Table 5 and Table 6. The width of the bid-ask spread first decreases then increases as the RRA increases, and the option spreads get wider from ITM to far OTM options for both depths of quotes. Before proceeding any further with empirical results, we need to explain certain apparently implausible properties of option bounds discussed above. It is shown in Table 5 that the proportional spread is not monotonically increasing in the RRA, as one would expect. This is in apparent contradiction of the expected relation of the option bounds to RRA, namely, we expect the lower bound to increase and the upper bound to decrease as the RRA increases.

We explain the observed non-monotonic relation by the properties of the NT region combined with the incompleteness of our model, which cannot estimate the optimal option bid and ask prices but only bounds on them, as discussed in Section 3.3. This incompleteness consists in the assumption that the NT region and the corresponding value function in the two-asset portfolio case remain valid in the presence of a short or long option position. Thus, the selected mid-point of this region is around where the value function reaches its maximum, since it is near this midpoint that the expected transactions costs reach their minimum. Also, note that in general the midpoint of this region is located near the Merton line, the optimal portfolio composition for the frictionless case. Figure 1 illustrates the value function for the above parameter set and an RRA of 4, plotted as a function of its location represented as a percentage of the NT region.

[Insert Figure 1 here]

Unreported results show that this function becomes steeper as the time horizon and RRA increases. Another set of unreported results shows that the derivative of the value function in the upper bound of the option price is lower for lower RRA, as expected. The explanation for our numerical results comes from the relative heights of the value function for different values of RRA, i.e., even though this function increases more per unit of the upper bound for a *lower* RRA, in some cases this bound may be higher for a higher RRA. In other words, in certain scenarios the disutility of getting away from the preferred habitat around the midpoint of the NT region is sufficiently high so as to prevail over the disutility of receiving a lower price for the short position (option upper bound).

We illustrate that indeed the height of the value function without options for long investment horizons is the source of the problem. For two values of RRA, δ of 2 and of 4, and for the depth of the quote n_c of 0.2, Table 7 displays the lower and upper bounds for call options for the overall investment horizon T of one and of ten years, with other parameter values as above. We clearly see from Table 7 that for the shorter horizon $T=1$ the option prices and proportional spreads are as expected, namely wider for $\delta=4$, whereas for the longer horizon $T=10$ these spreads unexpectedly tighten for OTM options. Unfortunately, a rigorous derivation of the tightest possible bounds for the two-asset portfolio in the presence of options does not exist in the literature and will not be attempted in this study; in fact, it has not even been shown that in such case the investment policy remains simple and the NT region is still a cone.

[Insert Table 7 here]

Tables 8 and 9 present the statistics of the proportional spread of our derived prices and observed market prices for both call and put options. At each observation date we select the observed market option prices whose moneyness values are closest to 0.95, 1, and 1.05; from our derived reservation prices as functions of RRAs we select the ones that best match those observed market prices. The results show that on average, our derived prices have a wider spread in comparison to the market prices, especially for OTM options.

[Insert Table 8 and Table 9 here]

5.2 Option Implied Risk Aversion

We estimate the level of risk aversion for each cross section in five different cases. In the base case, all the derived option reservation prices are used simultaneously to estimate the RRA as in (3.27)–(3.28). For the call only (put only) case we use the cross-section's call (put) bid-ask prices exclusively, and for the bid only (ask only) case, we use bid (ask) prices of both call and put options. Tables 10 and 11 display the statistics of the estimated RRA of market maker and investor, respectively.

[Insert Table 10 and Table 11 here]

The smaller depth of quote generates a higher RRA for both market maker and investor, suggesting that market maker and investor would trade with a lower volume if they tend to be more risk averse. The market maker's RRA is smaller than the investors', except for the bid only case, indicating that the market maker is less risk averse than investors in general and is more concerned about risks when purchasing options. These results are implied by the fact that the investor's upper bound (price to sell) needs to be lower than the market maker's lower bound (price to buy) for a trading to occur. The converse holds for the investor's lower bound, i.e., it needs to be higher than the market maker's upper bound.

In Figure 2, the 244 estimated RRAs of the market maker for the depth quote of 0.2 are plotted against observation dates. Similar results for the investor's RRA are shown in Figure 3. These figures show that the risk aversion coefficient becomes very high before the beginning of rare market events, and then drops significantly to a local minimum at the time when the market event starts. A possible interpretation is that during the market events the estimated volatility is high since the VIX is very high while the adjustment from VIX to physical volatility is roughly constant, and a relatively low RRA is sufficient to match the observed market prices. We also see in the figures that the RRA unexpectedly jumps to its upper limit at times when both market volatility and VIX are high. This is likely a result of the non-monotonicity of the spreads in RRA for our model, which implies that relatively low spreads observed at low volatility are best matched by high RRA.

[Insert Figure 2 and Figure 3 here]

Since a high probability of a declining market corresponds to a higher volatility, as shown in Figure 4 and Figure 5, the S&P 500 index price dropped significantly during the market events as the estimated RRA decreased. Market makers may suspect an impending financial crisis from the decreasing index price and increasing volatility, and therefore they are more reluctant to trade before the crisis bursts (regression results of market maker's RRA against S&P 500 prices and P-distributed volatility are presented in the Appendix).

[Insert Figure 4 and Figure 5 here]

The estimated RRA's of market maker and investor are regressed against three implied volatility (IV) smile characteristics: ATM implied volatility (ATM IV), the left skew (LS), and the right skew (RS). The ATM IV is the average implied volatility of the options whose call and put strike prices have their moneyness closest to 1. The left skew is the IV of the option whose put strike price has moneyness closest to 0.95 minus the ATM IV, and the right skew is the IV of the option whose call strike price has moneyness closest to 1.05 minus the ATM IV. Note that the ATM IV is generally very close to VIX. The regression model of the estimated RRA against the IV smile characteristics of market maker and of investor respectively are:

$$\begin{aligned}\hat{\delta}_{Mm} &= \alpha + \beta_1 ATMIV_t + \beta_2 LS_t + \beta_3 RS_t + \varepsilon_t \\ &. \\ \hat{\delta}_{Inv} &= \alpha + \beta_1 ATMIV_t + \beta_2 LS_t + \beta_3 RS_t + \varepsilon_t\end{aligned} \tag{5.1}$$

The regression results in Table 12 show that the estimated RRAs of market maker and investor are negatively related to the ATM IV. This negative relation is closely related to the previous observations about the relation of the spreads to RRA.

[Insert Table 12 here]

6. <u>Conclusions</u>

This study derived the level of risk aversion implied by the S&P 500 index option market over the time period of 1996 to 2016, under the market maker's two-asset portfolio selection and option pricing problems in the presence of proportional transaction costs in trading the underlying asset. The discretization of asset dynamics and the utility maximization approach were used under the assumption that the market maker has a finite investment horizon which exceeds the option maturity, following the methodology of Czerwonko and Perrakis (2011, 2016a, 2016b).

This is the first empirical study that recognizes the existence of proportional transactions costs in the problem of market making in the index option market. The applied model suggests that periods of high risk aversion implied by the option prices and their spreads precede adverse market events; we also observe a plausible result that the investor who wishes to trade with the market maker has a higher RRA than the market maker.

We also observe several counterintuitive results such as the non-monotonicity of the model-implied option spreads in RRA and the negative relation of RRA to the volatility level implied by the market option prices. This implies several issues in the modeling that need to be taken into account in future studies. Besides the properties of the value function related to the investment horizon mentioned in the preceding section, which cause non-monotonicity of the implied spreads in RRA, a problematic assumption used in this study is the re-derivation of the NT region at each new value of the volatility of the underlying index. If such a policy were applied in real life an investor would incur high transactions costs due to the frequent restructuring implied by the high sensitivity of the NT region to the volatility. The idea behind the optimality of the simple investment policy is to reduce the trading frequency. It is also likely that by assuming a constant portfolio policy under a constant RRA for the market maker we would observe a plausible behavior for the RRA implied by matching the observed option prices to the option bounds. In other words, a more plausible set of assumptions would use a constant volatility to derive the NT region, a constant RRA for the market maker, and a time-varying RRA for the investor.

References

Bensaid, B., Lesne, J., & Scheinkman, J. (1992). Derivative asset pricing with transaction costs. *Mathematical Finance, 2*(2), 63-86.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy, 81*(3), 637-654.

Boyle, P. P., & Lin, X. (1997). Optimal portfolio selection with transaction costs. *North American Actuarial Journal, 1*(2), 27-39.

Boyle, P. P., & Vorst, T. (1992). Option replication in discrete time with transaction costs. *The Journal of Finance, 47*(1), 271-293.

Constantinides, G. M. (1979). Multiperiod consumption and investment behavior with convex transactions costs. *Management Science, 25*(11), 1127-1137.

Constantinides, G. M. (1986). Capital market equilibrium with transaction costs. *Journal of Political Economy, 94*(4), 842-862.

Constantinides, G. M., & Perrakis, S. (2002). Stochastic dominance bounds on derivatives prices in a multiperiod economy with proportional transaction costs. *Journal of Economic Dynamics and Control, 26*(7), 1323-1352.

Constantinides, G. M., & Perrakis, S. (2007). Stochastic dominance bounds on american option prices in markets with frictions. *Review of Finance*, *11*(1), 71-115.

Constantinides, G. M., & Zariphopoulou, T. (1999). Bounds on prices of contingent claims in an intertemporal economy with proportional transaction costs and general preferences. *Finance and Stochastics, 3*(3), 345-369.

Constantinides, G. M., & Zariphopoulou, T. (2001). Bounds on derivative prices in an intertemporal setting with proportional transaction costs and multiple securities. *Mathematical Finance, 11*(3), 331-346.

Cox, J. C., Ross, S. A., & Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of Financial Economics, 7*(3), 229-263.

Czerwonko, M., & Perrakis, S. (2011). Transaction costs and dynamics discretization, with applications to asset allocation and option market making. *Working Paper.*

Czerwonko, M., & Perrakis, S. (2016a). Portfolio selection with transaction costs and jump-diffusion asset dynamics I: A numerical solution (January 10, 2016). *Forthcoming in Quarterly Journal of Finance,* Available at SSRN: https://ssrn.com/abstract=2897073

Czerwonko, M., & Perrakis, S. (2016b). Portfolio selection with transaction costs and jump-diffusion asset dynamics II: Economic implications. *Quarterly Journal of Finance, 6*(04), 1650019.

Davis, M. H., & Clark, J. M. C. (1994). A note on super-replicating strategies. *Philosophical Transactions: Physical Sciences and Engineering,* 485-494.

Davis, M. H., & Norman, A. R. (1990). Portfolio selection with transaction costs. *Mathematics of Operations Research, 15*(4), 676-713.

Davis, M. H., Panas, V. G., & Zariphopoulou, T. (1993). European option pricing with transaction costs. *SIAM Journal on Control and Optimization, 31*(2), 470-493.

Dumas, B., & Luciano, E. (1991). An exact solution to a dynamic portfolio choice problem under transactions costs. *The Journal of Finance, 46*(2), 577-595.

Gennotte, G., & Jung, A. (1994). Investment strategies under transaction costs: The finite horizon case. *Management Science, 40*(3), 385-404.

Hodges, S.D., Neuberger, A. (1989). Optimal replication of contingent claims under transaction costs. *Review of Futures Markets 8,* 222-239.

Kamrad, B., & Ritchken, P. (1991). Multinomial approximating models for options with k state variables. *Management Science, 37*(12), 1640-1652.

Leland, H. E. (1985). Option pricing and replication with transactions costs. *The Journal of Finance, 40*(5), 1283-1301.

Liu, H., & Loewenstein, M. (2002). Optimal portfolio selection with transaction costs and finite horizons. *The Review of Financial Studies, 15*(3), 805-835.

Magill, M. J., & Constantinides, G. M. (1976). Portfolio selection with transactions costs. *Journal of Economic Theory, 13*(2), 245-263.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics,* 247-257.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science,* 141-183.

Merton, R. C. (1989). On the application of the continuous-time theory of finance to financial intermediation and insurance. *The Geneva Risk and Insurance Review, 14*(3), 225-261.

Perrakis, S., & Lefoll, J. (1997). Derivative asset pricing with transaction costs: An extension. *Computational Economics, 10*(4), 359-376.

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance, 52*(1), 35-55.

Soner, H. M., Shreve, S. E., & Cvitanić, J. (1995). There is no nontrivial hedging portfolio for option pricing with transaction costs. *The Annals of Applied Probability,* 327-355.

Zakamouline, V. I. (2006). European option pricing and hedging with both fixed and proportional transaction costs. *Journal of Economic Dynamics and Control, 30*(1), 1-25.

Tables and Figures

Tables

Table 1: Prediction Error of Monthly Volatility (1996–2016)

The table presents the statistics of the prediction error of different methods. The prediction errors are calculated as the difference between the monthly estimated volatility predicted by a given method and the realized volatility. The Adjusted AFE model is the VIX on the observation date adjusted by the average forecast error cumulated from January 1990 to the given date. The Regression I model is estimated from the regression of the logarithm of realized volatility on the logarithm of VIX, with an intercept. The Regression II model is estimated from the same regression but without the intercept.

Prediction Model	Mean	Median	St. Dev.	Skew	Ex. Kurt.
Adjusted AFE	-0.0099	-0.0049	0.0479	-1.4990	5.7158
Regression I	-0.0154	-0.0065	0.0499	-2.4365	8.8690
Regression II	-0.0108	-0.0050	0.0482	-1.8215	6.8042

Table 2: Descriptive Statistics of Variables

This table presents the statistics of variables including volatility, interest rate, dividend and price of S&P 500 index in Panel A, as well as the statistics of moneyness, strike price, highest closing bid and lowest closing ask of S&P 500 index option in Panel B.

Panel A				
	Volatility	Interest Rate	Dividend	S&P 500 Index Price
Mean	0.1606	0.0232	0.0015	1276.00
Minimum	0.0541	0.0000	0.0006	606.37
1st Percentile	0.0551	0.0001	0.0007	643.00
5th Percentile	0.0664	0.0002	0.0009	749.50
25th Percentile	0.1031	0.0010	0.0012	1076.80
Median	0.1442	0.0167	0.0016	1241.00
75th Percentile	0.1891	0.0481	0.0018	1434.80
95th Percentile	0.3117	0.0543	0.0021	2028.70
99th Percentile	0.4703	0.0617	0.0028	2114.60
Maximum	0.7687	0.0626	0.0038	2125.80
Std. Dev	0.0894	0.0218	0.0004	350.4508
Skewness	2.6749	0.3049	0.8940	0.6568
Kurtosis	15.1409	1.4159	6.3915	3.1846
Observation	244	244	244	244

Panel B								
	Call				Put			
	Moneyness	Highest Closing Bid	Lowest Closing Ask	Strike Price	Moneyness	Highest Closing Bid	Lowest Closing Ask	Strike Price
Mean	1.015	21.25	22.67	1412	0.901	11.11	12.20	1287.8
Minimum	0.960	0.150	0.200	585.0	0.335	0.150	0.200	300.0
1st PCTL	0.961	0.150	0.335	660.0	0.646	0.150	0.250	588.3
5th PCTL	0.966	0.300	0.650	830.0	0.726	0.200	0.400	715.0
25th PCTL	0.988	3.319	4.100	1155	0.839	0.750	1.300	1020
Median	1.014	14.70	16.10	1345	0.916	3.375	4.200	1240
75th PCTL	1.042	34.90	37.00	1623.8	0.980	15.60	17.00	1540
95th PCTL	1.069	61.80	64.55	2115	1.028	46.47	49.00	1975
99th PCTL	1.078	78.00	80.93	2210	1.037	66.43	69.77	2120
Maximum	1.080	98.90	101.3	2290	1.040	86.90	89.70	2210
Std. Dev	0.033	20.59	21.31	385.3	0.096	15.70	16.42	376.1
Skewness	0.111	0.957	0.926	0.414	-0.746	1.869	1.833	0.323
Kurtosis	1.909	3.076	3.000	2.452	3.310	6.149	5.992	2.430
Observation	5919	5919	5919	5919	13016	13016	13016	13016

Table 3: Fixed-Horizon NT Region for Different RRAs and Volatilities

This table presents the derived NT region of investment horizon of 10 years $[\underline{b}_T, \overline{b}_T]$, for 15 risk aversion coefficients (δ) and for 6 volatilities (V) that are respectively, the minimum, 25th percentile, median, mean, 75th percentile and maximum of P-distributed volatility in the data set. Other parameters used are: riskless rate of 2.32%, transaction costs rate k of 0.25%, and the risk premium of 4%.

δ	V=0.0541		V=0.1031		V=0.1442		V=0.1606		V=0.1891		V=0.7687	
	$\underline{b}_T$	$\overline{b}_T$	$\underline{b}_T$	$\overline{b}_T$	$\underline{b}_T$	$\overline{b}_T$	$\underline{b}_T$	$\overline{b}_T$	$\underline{b}_T$	$\overline{b}_T$	$\underline{b}_T$	$\overline{b}_T$
0.5	1.12	39.81	1.12	11.43	1.12	5.46	1.12	4.25	1.12	2.86	0.09	0.17
1.5	1.12	13.76	1.12	3.06	1.20	1.32	1.00	1.02	0.69	0.78	0.03	0.06
2.5	1.12	7.48	1.40	1.56	0.72	0.79	0.57	0.65	0.40	0.48	0.02	0.03
3.5	1.12	4.90	1.04	1.07	0.50	0.58	0.40	0.47	0.28	0.35	0.01	0.02
4.5	1.12	3.58	0.79	0.85	0.38	0.45	0.30	0.37	0.21	0.27	0.01	0.02
5.5	1.12	2.80	0.64	0.70	0.31	0.37	0.25	0.31	0.17	0.22	0.01	0.02
6.5	1.12	2.29	0.54	0.60	0.26	0.32	0.21	0.26	0.15	0.19	0.01	0.01
7.5	1.12	1.93	0.46	0.52	0.23	0.28	0.18	0.23	0.13	0.16	0.01	0.01
8.5	1.14	1.67	0.41	0.46	0.20	0.24	0.16	0.20	0.11	0.14	0.01	0.01
9.5	1.36	1.46	0.36	0.42	0.18	0.22	0.14	0.18	0.10	0.13	0.00	0.01
10.5	1.24	1.31	0.33	0.38	0.16	0.20	0.13	0.16	0.09	0.12	0.00	0.01
15.5	0.85	0.88	0.22	0.26	0.11	0.14	0.09	0.11	0.06	0.08	0.00	0.01
20.5	0.63	0.67	0.16	0.20	0.08	0.10	0.07	0.08	0.05	0.06	0.00	0.00
25.5	0.51	0.55	0.13	0.16	0.07	0.08	0.05	0.06	0.04	0.05	0.00	0.00
30.5	0.42	0.46	0.11	0.13	0.05	0.07	0.04	0.05	0.03	0.04	0.00	0.00

Table 4: Mid-point of NT region at Option Maturity for Different RRAs and Volatilities

This table presents the midpoint of derived NT region $b^0_{T'}$ $(\equiv \frac{\underline{b}_{T'} + \overline{b}_{T'}}{2})$ at option maturity T' for 15 risk aversion coefficients (δ) and for 6 volatilities (V) that are respectively, the minimum, 25th percentile, median, mean, 75th percentile and maximum of P-distributed volatility in the data set. Other parameters used are: riskless rate of 2.32%, transaction costs rate k of 0.25%, and the risk premium of 4%.

δ	V=0.0541	V=0.1031	V=0.1442	V=0.1606	V=0.1891	V=0.7687
	$b^0_{T'}$	$b^0_{T'}$	$b^0_{T'}$	$b^0_{T'}$	$b^0_{T'}$	$b^0_{T'}$
0.5	20.46	6.27	3.29	2.68	1.99	0.13
1.5	7.43	2.09	1.26	1.01	0.73	0.04
2.5	4.29	1.48	0.75	0.61	0.44	0.03
3.5	3.01	1.05	0.54	0.43	0.31	0.02
4.5	2.35	0.82	0.42	0.34	0.24	0.01
5.5	1.96	0.67	0.34	0.28	0.20	0.01
6.5	1.70	0.57	0.29	0.23	0.17	0.01
7.5	1.52	0.49	0.25	0.20	0.15	0.01
8.5	1.40	0.43	0.22	0.18	0.13	0.01
9.5	1.41	0.39	0.20	0.16	0.11	0.01
10.5	1.28	0.35	0.18	0.14	0.10	0.01
15.5	0.86	0.24	0.12	0.10	0.07	0.00
20.5	0.65	0.18	0.09	0.07	0.05	0.00
25.5	0.53	0.14	0.07	0.06	0.04	0.00
30.5	0.44	0.12	0.06	0.05	0.03	0.00

Table 5: Proportional Spread for the Call Option

This table presents the derived call option prices in the form of proportional spread, which is calculated as $\frac{2(c_a - c_b)}{(c_a + c_b)}$, for 15 different RRAs and for both depth of the quote n_c of 0.2 and of 0.1. Other parameters used are: riskless rate of 2.32%, estimated volatility of 16.06%, dividend yield of 0.15%, transaction costs rate k of 0.25%, and the risk premium of 4%. Investment horizon is 10 years.

	$n_c = 0.2$				$n_c = 0.1$			
δ	K/S=0.975	K/S=1	K/S=1.025	K/S=1.05	K/S=0.975	K/S=1	K/S=1.025	K/S=1.05
0.5	12.86	20.70	38.93	90.51	21.40	36.69	72.82	175.60
1.5	6.58	10.97	21.25	50.71	11.04	19.71	40.36	99.42
2.5	5.98	9.70	18.09	42.06	9.23	16.36	33.19	81.32
3.5	6.30	9.77	17.32	38.91	8.78	15.27	30.45	73.88
4.5	6.81	10.18	17.27	37.52	8.69	14.85	29.12	69.94
5.5	7.40	10.74	17.54	36.91	8.77	14.74	28.42	67.58
6.5	8.04	11.38	17.98	36.74	8.94	14.78	28.07	66.06
7.5	8.70	12.08	18.52	36.81	9.17	14.93	27.92	65.06
8.5	9.38	12.80	19.12	37.06	9.42	15.14	27.90	64.38
9.5	10.07	13.55	19.77	37.41	9.70	15.39	27.98	63.92
10.5	10.77	14.31	20.46	37.84	10.00	15.68	28.12	63.63
15.5	14.33	18.23	24.13	40.65	11.62	17.35	29.33	63.48
20.5	17.94	22.24	27.99	43.95	13.35	19.20	30.94	64.34
25.5	21.55	26.27	31.92	47.42	15.12	21.14	32.72	65.59
30.5	25.15	30.31	35.87	50.99	16.91	23.11	34.58	67.05

Table 6: Proportional Spread for the Put Option

This table presents the derived put option prices in the form of proportional spread, which is calculated as $\frac{2(p_a - p_b)}{(p_a + p_b)}$, for 15 different RRAs and for both depth the quote n_c of 0.2 and of 0.1. Other parameters used are: riskless rate of 2.32%, estimated volatility of 16.06%, dividend yield of 0.15%, transaction costs rate k of 0.25%, and the risk premium of 4%. Investment horizon is 10 years.

	$n_c = 0.2$				$n_c = 0.1$			
δ	K/S=0.95	K/S=0.975	K/S=1	K/S=1.025	K/S=0.95	K/S=0.975	K/S=1	K/S=1.025
0.5	100.05	37.49	18.14	10.81	198.34	72.92	33.96	19.16
1.5	60.83	22.89	11.12	6.54	119.66	43.61	20.03	11.00
2.5	50.27	19.37	9.77	5.89	97.93	35.86	16.62	9.19
3.5	46.17	18.39	9.73	6.15	88.76	32.80	15.45	8.70
4.5	44.20	18.16	10.05	6.59	83.84	31.27	14.97	8.59
5.5	43.18	18.29	10.52	7.12	80.83	30.43	14.80	8.63
6.5	42.67	18.60	11.08	7.69	78.85	29.97	14.80	8.77
7.5	42.48	19.03	11.69	8.29	77.49	29.72	14.90	8.96
8.5	42.48	19.52	12.32	8.90	76.52	29.62	15.07	9.19
9.5	42.62	20.07	12.98	9.53	75.83	29.62	15.27	9.44
10.5	42.86	20.65	13.66	10.17	75.33	29.69	15.51	9.70
15.5	44.82	23.82	17.16	13.41	74.43	30.60	16.96	11.16
20.5	47.39	27.20	20.76	16.69	74.75	31.95	18.61	12.73
25.5	50.18	30.67	24.38	19.98	75.55	33.48	20.34	14.34
30.5	53.08	34.16	28.00	23.27	76.62	35.09	22.10	15.96

Table 7: Call Option Bounds for Investment Horizon of One and of Ten Years

This table presents the call lower and upper bounds for the overall investment horizon T of one year and of ten years and for the depth the quote n_c of 0.2, using the risk aversion coefficient (δ) of 2 and 4. The proportional spread is calculated as $\frac{2(c_a - c_b)}{(c_a + c_b)}$. Other parameters used are: riskless rate of 2.32%, estimated volatility of 16.06%, dividend yield of 0.15%, transaction costs rate k of 0.25%, and the risk premium of 4%.

	δ=2			δ=4		
K/S	Bid	Ask	Spread (%)	Bid	Ask	Spread (%)
T=1 year						
0.95	5.23	5.45	4.02	5.21	5.46	4.71
0.975	3.25	3.45	5.91	3.24	3.45	6.50
1	1.73	1.91	9.82	1.73	1.91	9.87
1.025	0.79	0.94	17.06	0.79	0.95	18.75
1.05	0.28	0.41	37.52	0.27	0.43	44.22
T=10 years						
0.95	5.20	5.42	4.08	5.18	5.44	4.73
0.975	3.22	3.42	5.99	3.21	3.43	6.54
1	1.71	1.89	9.97	1.71	1.88	9.95
1.025	0.78	0.94	19.08	0.78	0.93	17.24
1.05	0.27	0.42	45.12	0.28	0.41	38.07

Table 8: Statistics of Estimated Spread and Market Spread for Call Option

This table presents statistics of derived spread and market spread of the call option, and the proportional spread is calculated as $\frac{2(c_a - c_b)}{(c_a + c_b)}$. The market call option prices are found with moneyness nearest to 0.95, 1, and 1.05 at each observation date. The derived call option prices that best match with the market prices are selected.

Spread (%)	K/S=0.95		K/S=1		K/S=1.05	
	Market	Derived	Market	Derived	Market	Derived
Mean	3.86	12.44	6.61	22.40	28.21	86.86
Min	1.58	2.45	1.54	-2.88	1.40	7.25
5^{th} PCTL	2.46	7.06	3.09	8.75	5.83	17.97
25^{th} PCTL	3.07	10.21	4.89	15.62	11.76	34.79
Median	3.61	12.42	6.38	20.89	1.18	61.40
75^{th} PCTL	4.28	14.87	8.12	25.48	40.00	135.36
95^{th} PCTL	6.34	17.29	10.82	46.02	76.58	199.89
99^{th} PCTL	8.29	19.57	15.07	57.53	109.55	199.93
Max	8.86	21.52	16.39	59.91	140.00	199.93
Mode	2.92	2.45	5.90	-2.88	46.15	7.25
Std. Dev	1.21	3.11	2.43	10.57	23.63	65.87
Observation	244	244	244	244	244	244

Table 9: Statistics of Estimated Spread and Market Spread for Put Option

This table presents statistics of derived spread and market spread of the put option, and the proportional spread is calculated as $\frac{2(p_a - p_b)}{(p_a + p_b)}$. The market put option prices are found with moneyness nearest to 0.95, 1, and 1.05 at each observation date. The derived put option prices that best match with the market prices are selected.

Spread (%)	K/S=0.95		K/S=1		K/S=1.05	
	Market	Derived	Market	Derived	Market	Derived
Mean	13.91	106.43	6.50	23.84	4.39	7.68
Min	2.02	-19.78	1.23	1.43	1.93	-1.68
5^{th} PCTL	5.35	20.69	3.12	9.32	2.84	3.76
25^{th} PCTL	9.05	41.72	4.95	14.74	3.58	5.52
Median	12.15	93.09	6.43	20.71	4.07	7.47
75^{th} PCTL	18.06	186.82	8.00	27.82	4.75	9.34
95^{th} PCTL	26.60	199.96	10.15	54.62	6.95	12.95
99^{th} PCTL	32.49	199.96	12.56	75.08	9.06	15.20
Max	34.48	199.97	14.12	82.21	10.22	17.14
Mode	18.18	-19.78	6.45	1.43	1.93	-1.68
Std. Dev	6.55	68.69	2.23	14.09	1.31	2.90
Observation	244	244	244	244	244	244

Table 10: Statistics of Estimated RRA of the Market Maker

This table displays the statistics of the estimated RRA of market maker $\hat{\delta}_{Mm}$ for the depth of the quote n_c = 0.2 and n_c = 0.1. There are 244 RRAs estimated from fitting the derived option prices with observed market prices. In the base case, the RRA is estimated using all the derived option prices, whereas in other four cases, namely call, put, bid, ask only case, the RRA is estimated exclusively using derived call bid-ask prices, put bid-ask prices, bid of call and put prices, ask of call and put prices, respectively. 244 corresponding riskless rates, estimated volatilities, and dividend yields are used. The transaction costs rate k is 0.25%, the risk premium is 4%, and the investment horizon is 10 years.

	Base Case		Call Only		Put Only		Bid Only		Ask Only	
n_c	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1
Mean	10.61	17.01	14.53	18.16	9.66	16.25	14.36	17.17	13.94	18.38
Minimum	0.50	0.50	0.50	1.48	0.50	0.50	0.81	0.92	0.50	0.50
25th PCTL	2.00	5.57	4.85	8.61	1.49	4.13	5.14	8.37	2.18	5.85
Median	6.24	16.49	10.19	17.09	3.60	12.27	10.42	15.35	9.10	21.45
75th PCTL	16.57	30.49	25.97	30.49	16.88	30.49	23.81	29.86	30.49	30.49
95th PCTL	30.50	30.50	30.50	30.50	30.50	30.50	30.50	30.50	30.50	30.50
Maximum	30.50	30.50	30.50	30.50	30.50	30.50	30.50	30.50	30.50	30.50
Mode	0.50	0.50	30.50	30.50	0.50	0.50	0.50	30.50	30.50	0.50
Std. Dev	10.37	11.87	10.84	10.54	10.97	12.58	10.43	10.04	12.43	12.20
Observation	244	244	244	244	244	244	244	244	244	244

Table 11: Statistics of Estimated RRA of the Investor

This table displays the statistics of the estimated RRA of investor $\widehat{\upsilon}_{Inv}$ for the depth of the quote n_c = 0.2 and n_c = 0.1. There are 244 RRAs estimated from fitting the derived option prices with observed market price reversely. In the base case, the RRA is estimated using all the derived option prices, whereas in other four cases, namely call, put, bid, ask only case, the RRA is estimated using derived call bid-ask prices, put bid-ask prices, bid of call and put prices, ask of call and put prices, respectively. 244 corresponding riskless rates, estimated volatilities, and dividend yields are used. The transaction costs rate k is 0.25%, the risk premium is 4%, and the investment horizon is 10 years.

	Base Case		Call Only		Put Only		Bid Only		Ask Only	
n_c	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1
Mean	15.98	22.74	14.64	19.11	15.17	21.15	12.68	17.34	17.62	23.23
Minimum	0.50	0.50	0.50	0.50	0.50	0.50	0.50	1.43	0.50	0.50
25^{th} PCTL	5.70	15.74	4.42	8.75	3.93	9.83	4.77	8.58	6.16	16.72
Median	15.50	30.49	10.36	20.39	10.49	30.49	8.58	15.51	18.71	30.49
75^{th} PCTL	29.77	30.49	30.34	30.49	30.49	30.49	19.65	29.48	30.49	30.49
95^{th} PCTL	30.50	30.50	30.50	30.50	30.50	30.50	30.50	30.50	30.50	30.50
Maximum	30.50	30.50	30.50	30.50	30.50	30.50	30.50	30.50	30.50	30.50
Mode	0.50	30.50	0.50	0.50	0.50	0.50	30.50	30.50	0.50	0.50
Std. Dev	11.18	10.61	11.36	10.85	12.14	11.24	9.61	9.94	11.64	10.61
Observation	244	244	244	244	244	244	244	244	244	244

Table 12: Summary of Regression results for RRA against Volatility Smile Characteristics

This table presents the results for the regression of RRA on volatility smile characteristics for both market maker and investor:

$$\hat{\delta}_{Mm} = \alpha + \beta_1 ATMIV_t + \beta_2 LS_t + \beta_3 RS_t + \varepsilon_t ,$$

$$\hat{\delta}_{Inv} = \alpha + \beta_1 ATMIV_t + \beta_2 LS_t + \beta_3 RS_t + \varepsilon_t$$

where α is the intercept, AIM IV is at-the-money implied volatility, LS is left skew and RS is right skew. Values presented in brackets are t-values.

Dependent Variable	$\hat{\delta}_{Mm}$		$\hat{\delta}_{Inv}$	
n_c	0.2	0.1	0.2	0.1
Intercept	26.18***	43.07***	41.06***	49.68***
	(7.76)	(11.36)	(12.74)	(17.08)
ATM IV	-0.48***	-0.61***	-0.72***	-0.79***
	(-5.38)	(-6.06)	(-8.39)	(-10.18)
LS	-1.19	-3.69***	-2.36***	-2.85***
	(-1.54)	(-4.24)	(-3.20)	(-4.27)
RS	0.64	-0.26	0.74	0.17
	(0.95)	(-0.34)	(1.15)	(0.29)
Observation	244	244	244	244

*** Significant at 99% Confidence level

** Significant at 95% Confidence level

* Significant at 90% Confidence level

Figures

Figure 1.

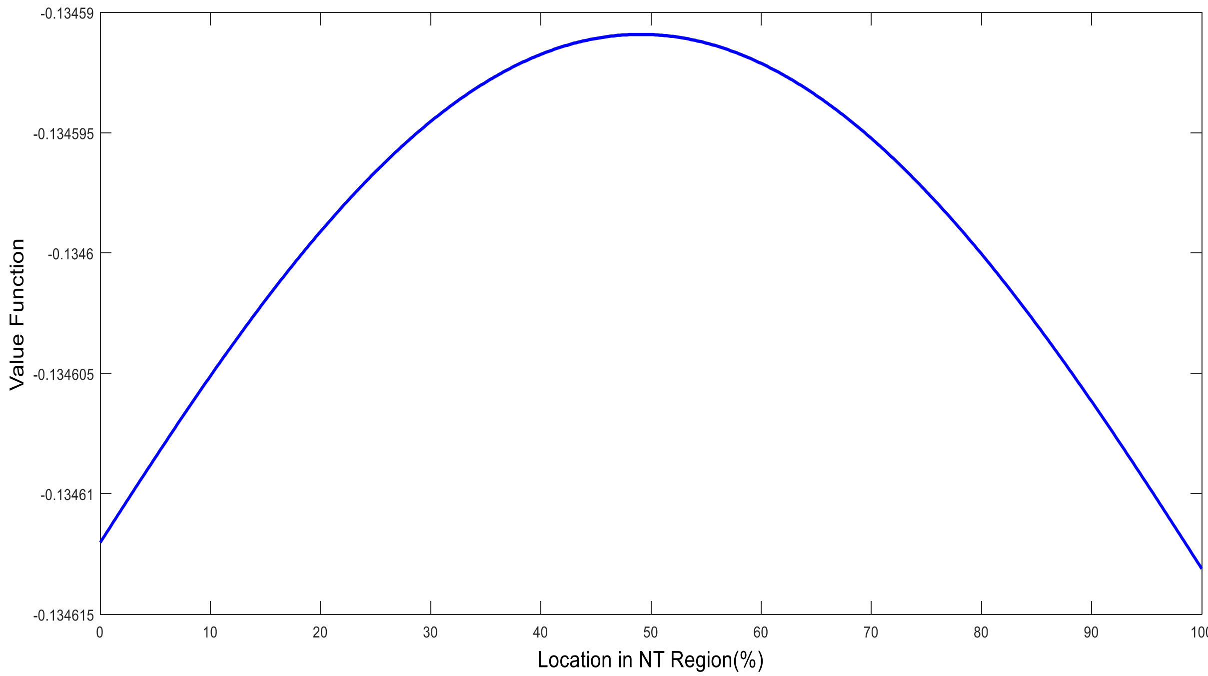

Figure 1: Value Functions for RRA of Four against the Percentage of the NT Region

Figure 2.

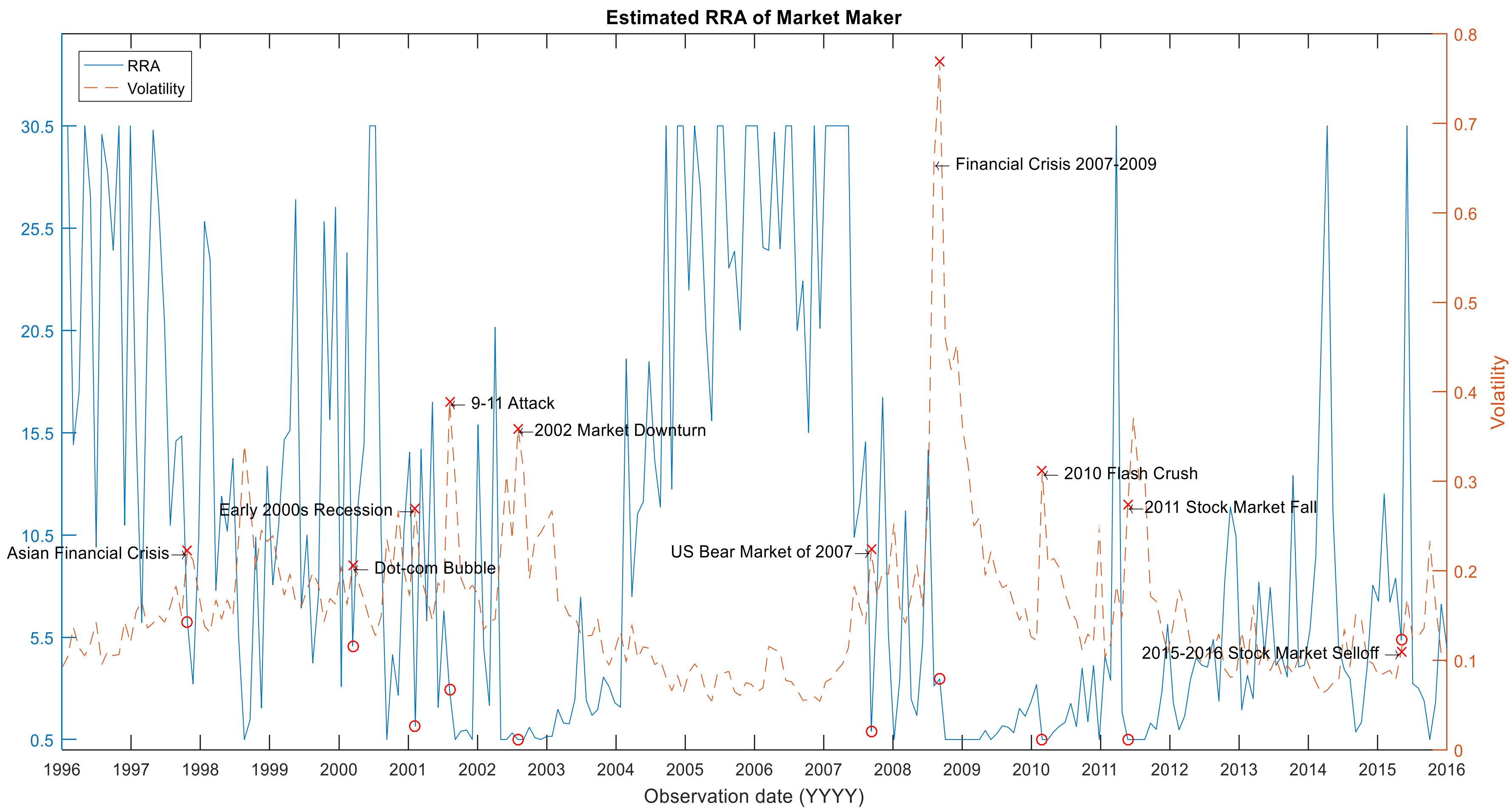

Figure 2: Estimated Risk Aversion of Market Maker and Volatility (1996–2016)

Figure 3.

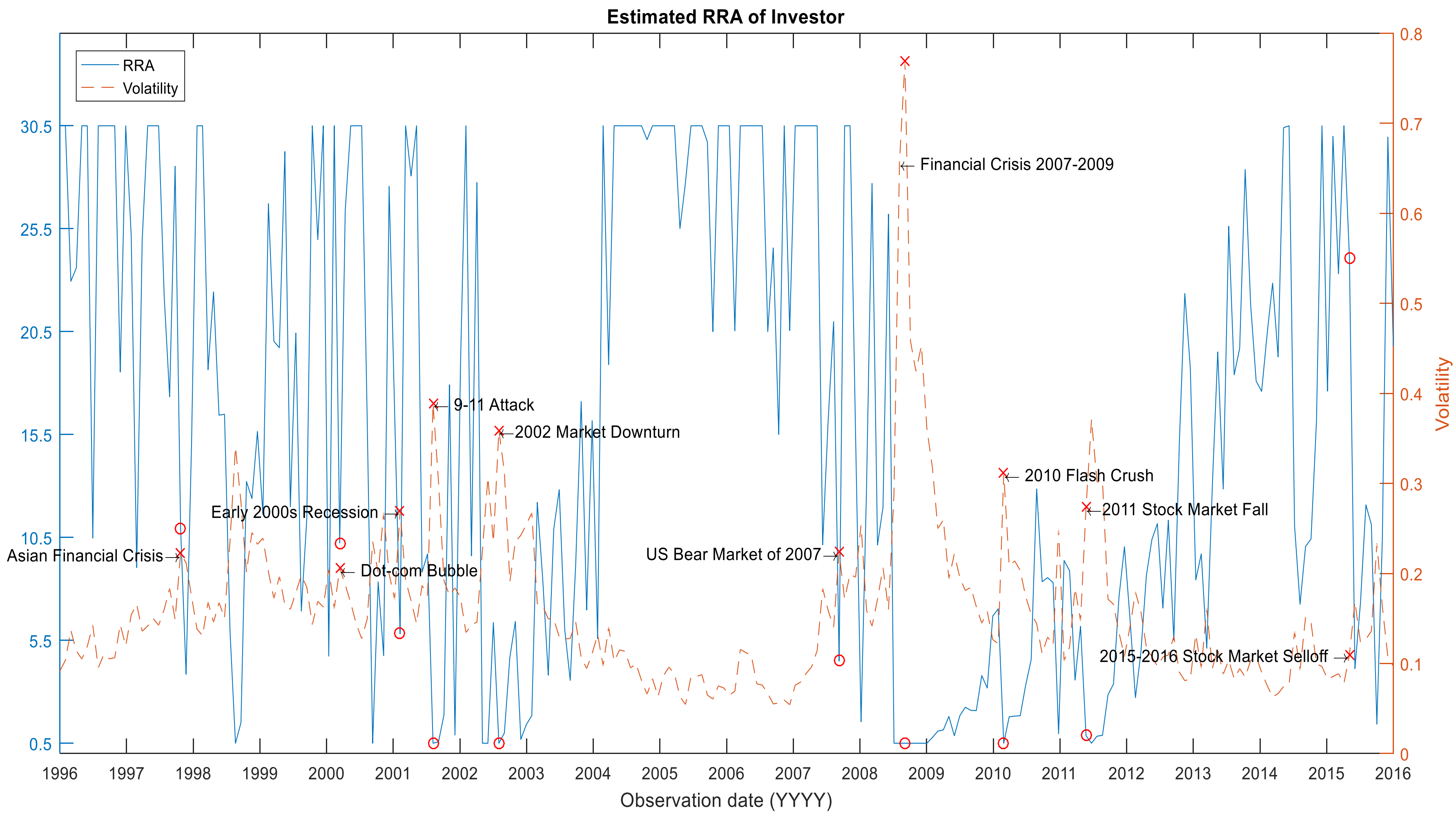

Figure 3: Estimated Risk Aversion of Investor and Volatility (1996–2016)

Figure 4.

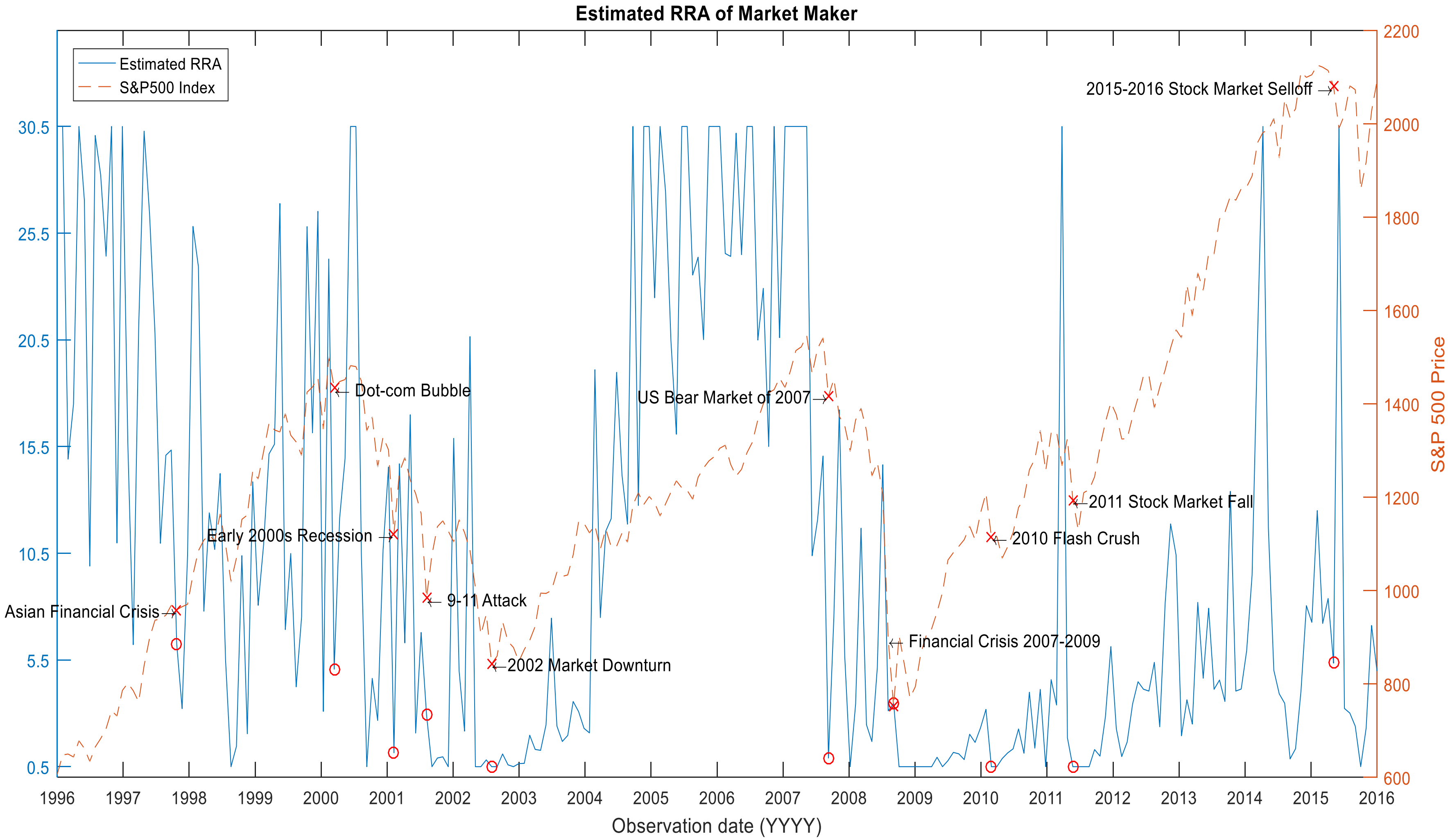

Figure 4: Estimated Risk Aversion of Market Maker and S&P 500 Index Price (1996–2016)

Figure 5:

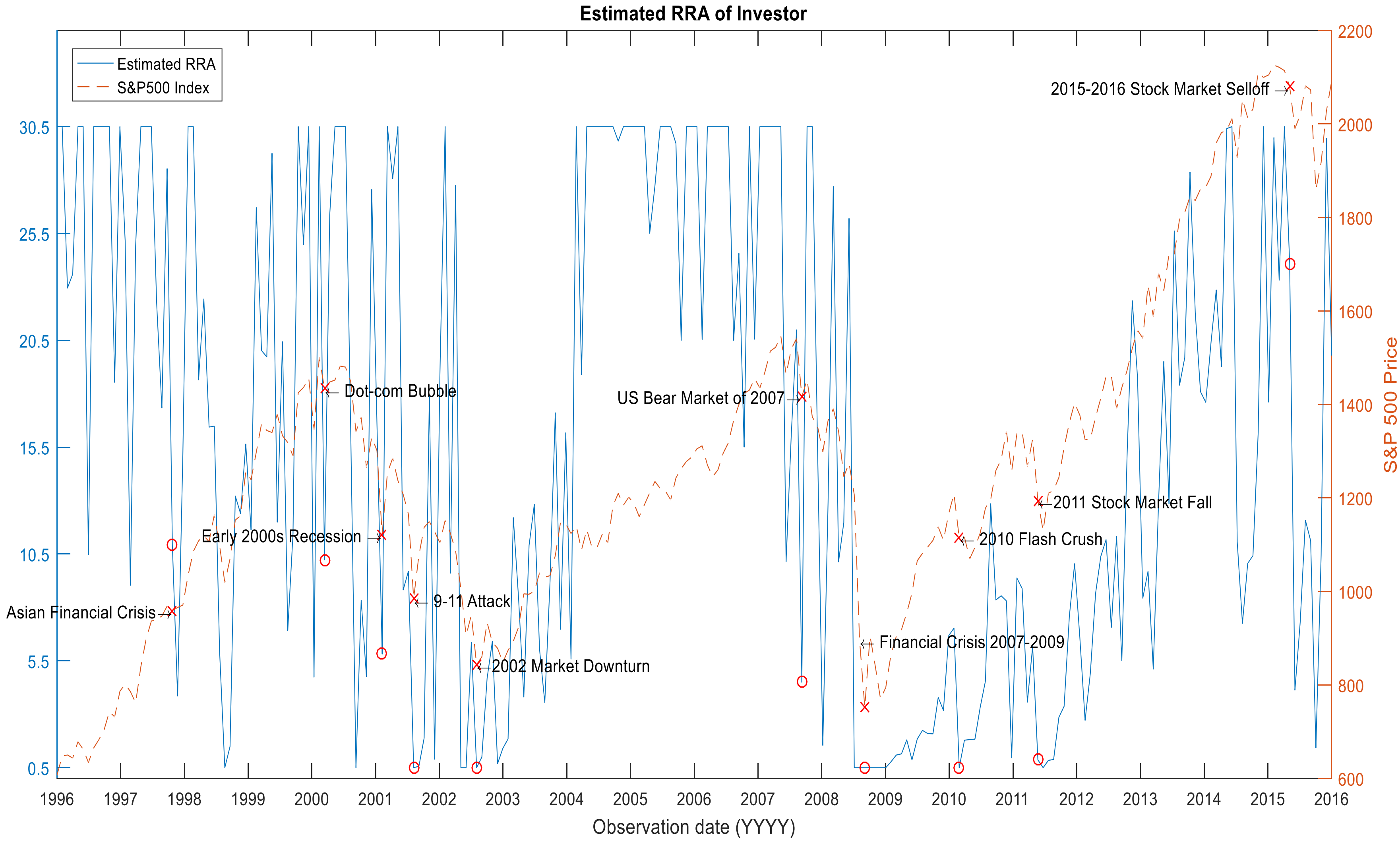

Figure 5: Estimated Risk Aversion of Investor and S&P 500 Index Price (1996–2016)

Appendix

Table A1: List of Adverse Market Events in the United States (1996–2016)

This table presents 10 major adverse market events in the United States between 1996 to 2016.

Time (YYYY-MM)	Market Event
1997-10	Asian Financial Crisis affected U.S.
2000-03	Dot-com Bubble
2001-03	Early 2000s Recessions
2001-09	9-11 Attacks
2002-09	Stock Market Downturn
2007-10	U.S. Bear Market
2008-10	Financial Crisis 2008
2010-05	Flash Crush (U.S.)
2011-08	Stock Market Fall
2015-08	U.S. Stock Market Sell-off

Table A2: Regression results for RRA of Market Maker against S&P 500 Prices and Volatility

This table presents the results for the regression of market maker's RRA on S&P 500 Prices and Volatility:

$$\hat{\delta}_{Mm} = \alpha + \beta_1 S_t + \beta_2 Vol_t + \varepsilon_t ,$$

where α is the intercept, S_t is the S&P 500 price, and Vol_t is the P-distributed volatility. Values presented in brackets are t-values.

Dependent Variable	$\hat{\delta}_{Mm}$	
n_c	0.2	0.1
Intercept	30.63***	40.36***
	(10.49)	(12.05)
S&P 500 prices	–0.0077***	–0.0093***
	(–4.36)	(–4.56)
Volatility	–63.39***	–71.83***
	(–9.15)	(–9.04)
Observation	244	244

*** Significant at 99% Confidence level
** Significant at 95% Confidence level
* Significant at 90% Confidence level